Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Filer:	Thomas C. Meyers/Sandra Richards
Attorney Docket Number:	NUGN-001/03US 34079/516

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
UTILITY APPLICATION FILING	1011	1	300	300
UTILITY SEARCH FEE	1111	1	660	660
UTILITY EXAMINATION FEE	1311	1	760	760
Pages:				
Claims:				
CLAIMS IN EXCESS OF 20	1202	1	100	100
Miscellaneous-Filing:				
Petition:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**1820**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	NUGEN TECHNOLOGIES, INC.
SERIAL NUMBER:	NOT YET ASSIGNED
FILING DATE:	NOT YET ASSIGNED
EXAMINER:	NOT YET ASSIGNED
CONTINUITY:	Continuation of Application No. 15/406,002, filed January 13, 2017
TITLE:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

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

FILED ELECTRONICALLY

PRELIMINARY AMENDMENT

Prior to examination of the application, please enter this amendment.

Amendment to the specification begin on page 2.

Remarks begin on page 3.

Applicant authorizes any required fee to be charged to Deposit Account No. 500369.

Remarks

By this amendment, the specification is amended to state the full benefit claim.

No new matter is added.

If there are any questions concerning this application, please contact Applicant's undersigned representative.

Dated: January 15, 2020

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: 617-856-8483

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

63611281

IDT Exhibit 1019.0004

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

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
	Douglas	A.	Amorese	

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

City	State/Province	Country of Residence
Los Altos	CA	US

Mailing Address of Inventor:

Address 1	900 Chesapeake Drive
Address 2	
City	Redwood City
State/Province	CA
Postal Code	94063
Country	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Jonathan		Scolnick	

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

City	State/Province	Country of Residence
San Francisco	CA	US

Mailing Address of Inventor:

Address 1	451 Kansas Street #518
Address 2	
City	San Francisco
State/Province	CA
Postal Code	94107
Country	US

Inventor 3 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Benjamin	G.	Schroeder	

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

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

City	San Mateo	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	33 Hayward Ave, Apt 112		
Address 2			
City	San Mateo	**State/Province**	CA
Postal Code	94401	**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	21710		
Email Address	ip@brownrudnick.com	Add Email	Remove Email

Application Information:

Title of the Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		
Attorney Docket Number	NUGN-001/03US 34079/516	**Small Entity Status Claimed**	☐
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	6	**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.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

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	21710		

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	15/406002	2017-01-13

Prior Application Status	Patented				Remove
Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)	Patent Number	Issue Date (YYYY-MM-DD)
15/406002	Division of	14/540917	2014-11-13	9546399	2017-01-17

Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
14/540917	Claims benefit of provisional	61/903826	2013-11-13

Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button. Add

Foreign Priority Information:

IDT Exhibit 1019.0007

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

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.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

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 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

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	NUGEN TECHNOLOGIES, INC.		
Mailing Address Information For Applicant:			
Address 1	900 Chesapeake Drive		
Address 2			
City	Redwood City	**State/Province**	CA
Country	US	Postal Code	94063
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.

IDT Exhibit 1019.0010

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

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	/Thomas C. Meyers/			Date (YYYY-MM-DD)	2020-01-15
First Name	Thomas	Last Name	Meyers	Registration Number	36,989

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

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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	NUGN-001/03US 34079/516
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

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.

IDT Exhibit 1019.0013

CLAIMS

1. A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:
    a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:
        (i) an indexing primer binding site;
        (ii) an indexing site;
        (iii) an identifier site; and
        (iv) a target sequence primer binding site;
    b) amplifying the adapter-nucleic acid fragment ligated products;
    c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and
    d) detecting the population of sequencing reads comprising a sequencing read with a duplicate identifier site and target sequence.
2. The method of claim 1, wherein the method further comprises removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.
3. The method of claim 1, wherein the identifier site is sequenced with the indexing site.
4. The method of claim 1, wherein the identifier site is sequenced separately from the indexing site.
5. The method of claim 1, wherein the identifier site is sequenced with the target sequence.
6. The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.
7. The method of claim 1, wherein the adaptor comprises from 5' to 3':

(i) the indexing primer binding site;

(ii) the indexing site;

(iii) the identifier site; and

(iv) the target sequence primer binding site.

8. The method of claim 1, wherein the adaptor comprises from 5' to 3':

(i) the indexing primer binding site;

(ii) the indexing site;

(iii) the target sequence primer binding site; and

(iv) the identifier site.

9. The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.

10. The method of claim 9, wherein the nucleic acid fragments from each sample has the same indexing site.

11. The method of claim 10, wherein the sequencing reads are separated based on the indexing site.

12. The method of claim 11, wherein the separation of sequencing reads is performed prior to step d).

13. The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments.

14. The method of claim 13, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

15. The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length.

16. The method of claim 1, wherein the indexing site is about 6 nucleotides in length.

17. The method of claim 1, wherein the identifier site is between 1 and 8 nucleotides in length.

18. The method of claim 1, wherein the identifier site is about 8 nucleotides in length.

19. The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site.

20. The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.

21. A kit comprising a plurality of adaptors, wherein each adaptor comprises:

    (i) an indexing primer binding site;

    (ii) an indexing site;

    (iii) an identifier site; and

    (iv) a target sequencing primer binding site.

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Serial No. 61/903,826, filed November 13, 2013, which is herein incorporated by reference in its entirety.

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0002] The contents of the text file submitted electronically herewith are incorporated herein by reference in their entirety: A computer readable format copy of the Sequence Listing (filename: NUGN_001_01US_SeqList_ST25.txt, date recorded: November 12, 2014, file size 3 kilobytes).

FIELD OF THE PRESENT INVENTION

[0003] The present invention relates generally to the field of high throughput sequencing reactions and the ability to discern artifacts arising through sequence duplications from nucleotide molecules that are unique molecules.

BACKGROUND OF THE PRESENT INVENTION

[0004] In RNA sequencing applications, accurate gene expression measurements may be hampered by PCR duplicate artifacts that occur during library amplification. When analyzing RNA sequencing data, when two or more identical sequences are found, it can be difficult to know if these represent unique cDNA molecules derived independently from different RNA molecules, or if they are PCR duplicates derived from a single RNA molecule. In genotyping by

sequencing, duplicate reads can be considered non-informative and may be collapsed down to a single read, thus reducing the number of sequencing reads used in final analysis. Generally, sequencing reads may be determined to be duplicates if both forward and reverse reads have identical starting positions, even though two independently generated molecules can have identical starting positions by random chance. Single primer extension based targeted re-sequencing suffers from an issue in that only one end of a sequencing read is randomly generated, while the other (reverse read) end is generated by a specific probe. This may make it difficult to determine if two reads are duplicates because they have been duplicated by PCR or because by chance they happened to start at the same position.

[0005] In expression analysis studies there may be limited value in doing paired end sequencing since the goal of the experiment is to determine amounts of transcript present as opposed to studying exon usage. In these studies, paired end sequencing adds costs while the only value is in helping distinguish PCR duplicates. The probability of two reads starting in the same position on only one end is higher than the probability of two reads having the same starting position on two ends (forward and reverse read). There is a need for improved methods that allow for low-cost, high throughput sequencing of regions of interest, genotyping or simple detection of RNA transcripts without inherent instrument inefficiencies that drive up sequencing costs due to the generation of unusable or non-desired data reads. The invention described herein fulfills this need. Here, we describe an adaptor approach that allows for the identification of true PCR duplicates and their removal.

[0006] The methods of the present invention provide novel methods for identifying true duplicate reads during sequencing, such as to improve data analysis of sequencing data, and other related advantages.

SUMMARY OF THE PRESENT INVENTION

[0007] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The detection and/or removal of duplicate sequencing reads presented herein is a novel approach to increasing the efficacy of evaluating data generated from high throughput sequence reactions, including complex multiplex sequence reactions.

[0008] Accordingly, the present invention provides a method of detecting a duplicate sequencing read from a population of sample sequencing reads, the method comprising ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site. The ligated adaptor-nucleic acid fragment products can be amplified, thus generating a population of sequencing reads from the amplified adaptor-nucleic acid ligation products. The sequencing reads with a duplicate identifier site and target sequence can then be detected from the population of sequencing reads. The methods can further include the removal of the sequencing reads with the duplicate identifier site and target sequence from the population of sequence reads.

[0009] In some embodiments, the identifier site is sequenced with the indexing site or the target sequence. In further embodiments, the identifier site is sequenced separately from the indexing site or the target sequence.

[0010] In some embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the identifier site; and the target sequence primer binding site. In further embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the target sequence primer binding site; and the identifier site.

[0011] In some embodiments, the plurality of nucleic acid fragments is generated from more than one sample. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, the sequencing reads are separated based on the indexing site. In yet other embodiments, the separation of sequencing reads is performed prior to detecting sequence reads with a duplicate identifier site and target sequence.

[0012] In some embodiments, the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments. In further embodiments, the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

[0013] In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In further embodiments, the indexing site is about 6 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In further embodiments, the identifier site is about 8 nucleotides in length.

[0014] In some embodiments, the indexing primer binding site is a universal indexing primer binding site; and in some embodiments, the target sequence primer binding site is a universal target sequence primer binding site.

[0015] The present invention also encompasses embodiments that include a kit comprising a plurality of adaptors, wherein each adaptor comprise an indexing primer binding site; an indexing site, and identifier site, and a target sequencing primer binding site.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] A better understanding of the novel features of the invention and advantages of the present invention will be obtained by reference to the following description that sets forth

illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0017] Figure 1 depicts a schematic of generating sequencing reads of a library including where an indexing primer and a target sequence primer anneal.

[0018] Figure 2A depicts a mechanism of the Single Primer Enrichment Technology (SPET) and how an identifier site is carried over into the final library and how sequencing through the indexing site into the identifier site provides data on which nucleic acid molecule is being identified.

[0019] Figure 2B is a continuation of Figure 2A.

[0020] Figure 3 offers a detailed view of the sequence of an example adaptor, among many envisioned embodiments, and the position of the indexing and identifier sites (SEQ ID NO: 1). "N" refers to any nucleic acid.

[0021] Figure 4 depicts a schematic of two separate sequence libraries, indicating where the indexing primers and the target primers anneal in a traditional library as compared to a library using an identifier site.

[0022] Figure 5 depicts a data table demonstrating the accuracy of resolving true duplicates versus apparent or perceived duplicates using an identifier site in the adaptors.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0023] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The present invention encompasses methods of detecting sequences duplicated in sequencing applications, and further removal of the duplicated sequence reads. The present invention further encompasses kits

comprising components that would allow for customized applications of the method of detecting and removing duplicated sequence reads in high throughput sequencing reactions. The compositions and methods can be used with various applications for genetic sample analysis, such as RNA sequence analysis, copy number variation analysis, methylation sequencing analysis, genotyping and whole genome amplification.

[0024] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the present invention. On the contrary, the present disclosure is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the present invention.

[0025] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg 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.

[0026] In some embodiments, the methods disclosed herein is for detecting from a population of sequencing reads a sequencing read, such as a duplicate sequencing read with a duplicate identifier site and target sequence. A duplicate sequencing read can be a sequencing read with

the same identifier site and target sequence as another sequencing read in the population of sequencing reads.

Adaptor

[0027] The present invention provides compositions of adaptors and methods comprising use of an adaptor. An adapter refers to an oligonucleotide sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. The target polynucleotide molecules may be fragmented or not prior to the addition of adaptors. In some embodiments, a method disclosed herein comprises ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples.

[0028] Various adaptor designs are envisioned which are suitable for generation of amplification-ready products of target sequence regions/strands of interest. For example, the two strands of the adaptor may be self-complementary, non-complementary or partially complementary. In some embodiments, the adaptor can comprise an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site.

[0029] An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site. An indexing site is a nucleic acid sequence that acts as an index for multiple polynucleotide samples, thus allowing for the samples to be pooled together into a single sequencing run, which is known as multiplexing. In some embodiments, the indexing site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In some embodiments, the indexing site is about 6 nucleotides in length.

[0030] An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads. In some embodiments, the identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In some embodiments, the identifier site is about 8 nucleotides in length. This identifier site can be designed as a set of sequences, or it can be semi-random, or it can be completely random. In addition, this identifier site can be a fixed length, or it can be a variable length. In some embodiments, the identifier sites in a plurality of adaptors are of a fixed length. For example, the identifier sites can all be eight random bases. In another embodiment, the identifier sites in a plurality of adaptors are of a variable length. For example, the identifier sites can range in size from 1 to 8 bases. In yet another embodiment, the identifier sites can be of a defined set of defined sequence. For example, the identifier site of a plurality of adaptors can be one of 96 defined six-base nucleotide sequences.

[0031] A target sequence primer binding site nucleotide sequence for binding a primer for a target sequence. The primer can be used to amplify the target sequence (e.g., a nucleic acid fragment from a sample). Accordingly, in some embodiments, the adaptor comprises an indexing primer binding site and a target sequence primer binding site.

[0032] A primer is a polynucleotide chain, typically less than 200 residues long, most typically between 15 and 100 nucleotides long, but can encompass longer polynucleotide chains. A primer targeting the primer binding sites are typically designed to hybridize to single-stranded nucleic acid strands. In some embodiments, the primers targeting the primer binding sites are designed to hybridize to single-stranded DNA targets. In the case where the sample comprises genomic DNA or other double-stranded DNA, the sample can be first denatured to render the target single stranded and enable hybridization of the primers to the desired sequence regions of

interest. In these embodiments, the methods and compositions described herein can allow for region-specific enrichment and amplification of sequence regions of interest. In some embodiments, the other double-stranded DNA can be double-stranded cDNA generated by first and second strand synthesis of one or more target RNAs.

[0033] In other embodiments, the primers targeting the primer binding sites are designed to hybridize to double-stranded nucleic acid targets, without denaturation of the double stranded nucleic acids. In other embodiments, the primers targeting the primer binding sites are designed to hybridize to a double-stranded DNA target, without denaturation of the dsDNA. In these embodiments, the primers targeting the selected sequence regions of interest are designed to form a triple helix (triplex) at the selected sequence regions of interest. The hybridization of the primers to the double-stranded DNA sequence regions of interest can be carried out without prior denaturation of the double stranded nucleic acid sample. In such embodiments, the methods and compositions described herein can allow for region-specific enrichment as well as strand-specific enrichment and amplification of sequence regions of interest. This method can be useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and analysis of multiplicity of sequence regions of interest in the native complex nucleic acid sample. The method can find use for studies and analyses carried out in situ, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as permit the analysis of complex genomic DNA without disruption of chromatin structures.

[0034] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as for amplification of a target sequence (e.g., in PCR). The oligonucleotide primer can be a synthetic polynucleotide

that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a sequence of the target polynucleotide. In some embodiments, the 3' region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a primer binding site.

[0035] In some embodiments, the primer binding site is a binding site for a universal primer. A universal primer is a primer that can be used for amplifying a number of different sequences. In some embodiments, a universal primer is used to amplify different libraries. In some embodiments, an indexing primer binding site is a binding site for a universal indexing primer (i.e., the indexing primer binding site is a universal indexing primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal indexing primer binding site. In some embodiments, the universal indexing primer can be used to amplify and/or sequence a number of different indexing sites.

[0036] In some embodiments, a target sequence primer binding site is a binding site for a universal target sequence primer (i.e., the target sequence primer binding site is a universal target sequence primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal target sequence primer binding site. In some embodiments, the universal target sequence primer can be used to amplify and/or sequence a number of different target sequences.

[0037] In some embodiments, the adaptor comprises an identifier site 3' to an indexing site. In some embodiments, the adaptor comprises an identifier site 5' to an indexing site. In some embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target sequence primer binding site. In other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target

sequence primer binding site. In yet other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, identifier site, indexing site, and target sequence primer binding site.

Samples

[0038] In some embodiments, an adaptor is ligated to a nucleic acid fragment (e.g., the 5' end of the nucleic acid fragment). The nucleic acid fragment can be from a plurality of nucleic acid fragments from one or more samples. The nucleic acid fragment can be RNA, DNA, or complex DNA, for example genomic DNA and PNA, in which case one might use a modified nucleic acid. The nucleic acid fragment may also be cDNA. The cDNA can be generated from RNA, e.g., mRNA.

[0039] The sample can be a biological sample. For example, the sample can be an animal, plant, bacterial, algal, or viral sample. In some embodiments, the sample is a human, rat, or mouse sample. The sample can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The sample can be cDNA made from a mixture of genomes of different species. In some embodiments, the sample can be from a synthetic source. The sample can be mitochondrial DNA. The sample can be cell-free DNA. The cell-free DNA can be obtained from sources such as a serum or a plasma sample. The sample can comprise one or more chromosomes. For example, if the sample is from a human, the sample can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. In some embodiments, the sample comprises a linear or circular genome. The sample can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The sample can be from more than one individual or organism. The sample

can be double-stranded or single-stranded. The sample can be part of chromatin. The sample can be associated with histones.

[0040] In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from more than one sample, such as 2, 3, 4, 5, or more samples. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site.

[0041] In some embodiments, a population of nucleic acid fragments generated from a sample is of one or more specific size range(s). In some embodiments, the fragments have an average length from about 10 to about 10,000 nucleotides. In some embodiments, the fragments have an average length from about 50 to about 2,000 nucleotides. In some embodiments, the fragments have an average length from about 100-2,500, 10-1,000, 10-800, 10-500, 50-500, 50-250, or 50-150 nucleotides. In some embodiments, the fragments have an average length less than 10,000 nucleotide, such as less than 5,000 nucleotides, less than 2,500 nucleotides, less than 2,500 nucleotides, less than 1,000 nucleotides, less than 500 nucleotides, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides.

[0042] In some embodiments, fragmentation of the nucleic acids can be achieved through methods known in the art. Fragmentation can be achieved through physical fragmentation methods and/or enzymatic fragmentation methods. Physical fragmentation methods can include

nebulization, sonication, and/or hydrodynamic shearing. In some embodiments, the fragmentation can be accomplished mechanically comprising subjecting the nucleic acids in the input sample to acoustic sonication. In some embodiments, the fragmentation comprises treating the nucleic acids in the input sample with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of nucleic acid or polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g., from New England Biolabs). For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some embodiments, fragmentation comprises treating the nucleic acids in the input sample with one or more restriction endonucleases. Fragmentation can produce fragments having 5′ overhangs, 3′ overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel.

[0043] In some embodiments, fragmentation of the nucleic acids is followed by end repair of the nucleic acid fragments. End repair can include the generation of blunt ends, non-blunt ends (i.e sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3′-end of the nucleic acid fragments, by a polymerase lacking 3′-exonuclease activity. End repair can be performed using any number of enzymes and/or methods known in the art

including, but not limited to, commercially available kits such as the Encore™ Ultra Low Input NGS Library System I. In some embodiments, end repair can be performed on double stranded DNA fragments to produce blunt ends wherein the double stranded DNA fragments contain 5′ phosphates and 3′ hydroxyls. In some embodiments, the double-stranded DNA fragments can be blunt-end polished (or "end-repaired") to produce DNA fragments having blunt ends, prior to being joined to adapters. Generation of the blunt ends on the double stranded fragments can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded DNA fragments can be blunt ended by the use of a single stranded specific DNA endonuclease, for example, but not limited to, mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, or any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded fragments generated by fragmenting the sample comprising nucleic acids. In still other cases, the nucleic acid fragments can be made blunt ended by filling in the overhanging single stranded ends of the double stranded fragments. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions

of the double stranded fragments. Alternatively, the double stranded DNA fragments can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0044] U.S. Patent Publication Nos. 2013-0231253 A1 and 2014-0274729 A1 further describe methods of generating nucleic acid fragments, methods of modifying the fragments and analysis of the fragments, and are incorporated by reference in their entirety herein.

Ligation of Adaptors

[0045] Ligation of adaptors at the desired end of the sequence regions of interest (e.g., at the 5' or 3' end of a nucleic acid fragment generated form a sample) is suitable for carrying out the methods of the invention. Various ligation modalities are envisioned, dependent on the choice of nucleic acid, nucleic acid modifying enzymes and the resulting ligatable end of the nucleic acid. For example, when a blunt end product comprising the target region/sequence of interest is generated, blunt end ligation can be suitable. Alternatively, where the cleavage is carried out using a restriction enzyme of known sequence specificity, leading to the generation of cleavage sites with known sequence overhangs, suitable ends of the adaptors can be designed to enable hybridization of the adaptor to the cleavage site of the sequence region of interest and subsequent ligation. Ligation also refers to any joining of two nucleic acid molecules that results in a single nucleic acid sequence that can be further modified to obtain the sequence of the nucleic acids in question. Reagents and methods for efficient and rapid ligation of adaptors are commercially available, and are known in the art.

[0046] In some embodiments, the 5′ and/or 3′ end nucleotide sequences of fragmented nucleic acids are not modified or end-repaired prior to ligation with the adapter oligonucleotides of the present invention. For example, fragmentation by a restriction endonuclease can be used to leave

a predictable overhang, followed by ligation with one or more adapter oligonucleotides comprising an overhang complementary to the predictable overhang on a nucleic acid fragment. In another example, cleavage by an enzyme that leaves a predictable blunt end can be followed by ligation of blunt-ended nucleic acid fragments to adapter oligonucleotides comprising a blunt end. In some embodiments, end repair can be followed by an addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides, such as one or more adenine, one or more thymine, one or more guanine, or one or more cytosine, to produce an overhang. Nucleic acid fragments having an overhang can be joined to one or more adapter oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3′ ends of end-repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adapters each having a thymine at a 3′ end. In some embodiments, adapter oligonucleotides can be joined to blunt end double-stranded nucleic acid fragments which have been modified by extension of the 3′ end with one or more nucleotides followed by 5′ phosphorylation. In some cases, extension of the 3′ end can be performed with a polymerase such as for example Klenow polymerase or any of the suitable polymerases provided herein, or by use of a terminal deoxynucleotide transferase, in the presence of one or more dNTPs in a suitable buffer containing magnesium. In some embodiments, nucleic acid fragments having blunt ends can be joined to one or more adapters comprising a blunt end. Phosphorylation of 5′ ends of nucleic acid fragments can be performed for example with T4 polynucleotide kinase in a suitable buffer containing ATP and magnesium. The fragmented nucleic acid molecules may optionally be treated to dephosphorylate 5′ ends or 3′ ends, for example, by using enzymes known in the art, such as phosphatases.

[0047] In some embodiments, appending the adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction. In some embodiments, appendage of an adaptor to the nucleic acid fragments comprises ligation. In some embodiments, ligation of the adaptor to the nucleic acid fragments can be following end repair of the nucleic acid fragments. In another embodiment, the ligation of the adaptor to the nucleic acid fragments can be following generation of the nucleic acid fragments without end repair of the nucleic acid fragments. The adaptor can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors in which the adaptor comprises two complementary strands. In some embodiments, the adaptor can be a double stranded DNA adaptor. In some embodiments, the adaptor can be an oligonucleotide of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the adaptor is appended or attached. In some embodiments, the adaptor can be a conventional duplex adaptor, wherein the adaptor comprises sequence well known in the art. In a some embodiments, the adaptor can be appended to the nucleic acid fragments generated by the methods described herein in multiple orientations. In some embodiment, the methods described herein can involve the use of a duplex adaptor comprising double stranded DNA of known sequence that is blunt ended and can bind to the double stranded nucleic acid fragments generated by the methods described herein in one of two orientations. In some embodiments, the adaptor can be ligated to each of the nucleic acid fragments such that each of the nucleic acid fragments comprises the same adaptor. In other words, each of the nucleic acid fragments comprises a common adaptor. In another embodiment, an adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein such that each nucleic acid fragment in the library of nucleic acid

fragments comprises the adaptor ligated to one or both ends. In another embodiment, more than one adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein. The multiple adaptors may occur adjacent to one another, spaced intermittently, or at opposite ends of the nucleic acid fragments. In some embodiments, the adaptor can be ligated or appended to the 5′ and/or 3′ ends of the nucleic acid fragments generated by the methods described herein. The adaptor can comprise two strands wherein each strand comprises a free 3′ hydroxyl group but neither strand comprises a free 5′ phosphate. In some embodiments, the free 3′ hydroxyl group on each strand of the adaptor can be ligated to a free 5′ phosphate present on either end of the nucleic acid fragments of the present invention. In this embodiment, the adaptor comprises a ligation strand and a non-ligation strand whereby the ligation strand can be ligated to the 5′ phosphate on either end of the nucleic acid fragment while a nick or gap can be present between the non-ligation strand of the adaptor and the 3′ hydroxyl on either end of the nucleic acid fragment. In some embodiments, the nick or gap can be filled in by performing a gap repair reaction. In some embodiments, the gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In some embodiments, the gap repair can be performed using a DNA-dependent DNA polymerase with weak or no strand displacement activity. In some embodiments, the ligation strand of the adaptor can serve as the template for the gap repair or fill-in reaction. The gap repair or fill-in reaction may comprise an extension reaction wherein the ligation strand of the adaptor serves as a template and leads to the generation of nucleic acid fragments with complementary termini or ends. In some embodiments, the gap repair can be performed using Taq DNA polymerase. In some embodiments, the ligation of the first adaptor to the nucleic acid fragments generated by the

methods described herein may not be followed by gap repair. The nucleic acid fragments may comprise the adaptor sequence ligated only at the 5′ end of each strand.

[0048] Ligation and, optionally gap repair, of the adaptor to the nucleic acid fragments generates an adaptor-nucleic acid fragment complex. In some embodiments, the adaptor-nucleic acid fragment complex can be denatured. Denaturation can be achieved using any of the methods known in the art including, but not limited to, physical, thermal, and/or chemical denaturation. In some embodiments, denaturation can be achieved using thermal or heat denaturation. In some embodiments, denaturation of the adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising the adaptor sequence at only the 5′end of the nucleic acid fragments. In another embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising adaptor sequence at both the 5′end and 3′end of the nucleic acid fragments.

Methods of Amplification

[0049] The methods, compositions and kits described herein can be useful to generate amplification-ready products directly from a nucleic acid source for downstream applications such as next generation sequencing, as well as generation of libraries with enriched population of sequence regions of interest. In some embodiments, the adapter-nucleic fragment ligated products, e.g., from the ligation of an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, is amplified.

[0050] Methods of amplification are well known in the art. In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal. In

some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR).

[0051] Suitable amplification reactions can be exponential or isothermal and can include any DNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), linear amplification, multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No. 6,251,639), Ribo-SPIA, or a combination thereof. In some cases, the amplification methods for providing the template nucleic acid may be performed under limiting conditions such that only a few rounds of amplification (e.g., 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 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[0052] PCR is an in vitro amplification procedure based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5′ ends of the oligonucleotide primers.

[0053] LCR uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is

employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[0054] SDA (Westin et a 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3′-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[0055] Some aspects of the invention utilize linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

Downstream Applications

[0056] One aspect of the present invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as in next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. The methods of the present invention can also be used in the analysis of genetic information of selective genomic regions of interest (e.g., analysis of SNPs or other disease markers) as well as genomic regions which may interact with the selective region of interest. The methods of the present invention may further be used in the analysis of copy number variation as well as differential expression.

Sequencing

[0057] In some embodiments, a population of sequencing reads is generated from the amplified adapter-nucleic fragment ligated products. In some embodiments, a sequencing read comprises an index read, which comprises the sequence of an indexing site. In some embodiments, an index read comprises the sequence of an indexing site and the sequence of an identifier site. For example, the indexing site is sequenced with the identifier site. In some embodiments, the index read does not include the sequence of an identifier site. For example, the indexing site is not sequenced with the identifier site. In some embodiments, a sequencing read comprises the target sequence. In some embodiments, a sequencing read comprises a target sequence and an identifier sequence. For example, the target sequence is sequenced with the identifier site. In some embodiments, the target sequence is not sequenced with the identifier site.

[0058] The methods of the present invention are useful for sequencing by the method commercialized by Illumina, as described U.S. Pat Nos. 5,750,341; 6,306,597; and 5,969,119.

[0059] In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g., (A)/(A′) or both ends (e.g., (A)/(A′) and (C)/(C′)). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g., by SPIA or linear PCR).The resulting nucleic acid is then denatured and the single-stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time.

[0060] In some embodiments, the methods of the invention are useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature (*2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711;

US20060024678; US20080213770; and US20080103058. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452; 7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764.

[0061] An example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single head, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

[0062] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g., Soni G V and Meller A. (2007) *Clin Chem*

53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

Data Analysis

[0063] In some embodiments, the sequence reads are used in the analysis of genetic information of selective genomic regions of interest as well as genomic regions which may interact with the selective region of interest. Amplification methods as disclosed herein can be used in the devices, kits, and methods known to the art for genetic analysis, such as, but not limited to those found in U.S. Pat. Nos. 6,449,562, 6,287,766, 7,361,468, 7,414,117, 6,225,109, and 6,110,709.

[0064] In some embodiments, the sequencing reads are used to detect duplicate sequencing reads. In some embodiments, a sequencing read is identified as a duplicate sequencing read when it contains an identifier site and target sequence that is the same as another sequencing read from the same population of sequencing reads.

[0065] In some embodiments, duplicate sequencing reads are differentiated from one another as being true duplicates versus apparent or perceived duplicates. Apparent or perceived duplicates may be identified from sequencing libraries and using conventional measures of duplicate reads (i.e., reads were mapped using bowtie), where all reads with the same start and end nucleic acid coordinates were counted as duplicates. True duplicates may be identified from sequencing

libraries having had an identifier site introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates.

[0066] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acids generated from any dsDNA may have the same start site, as determined from the sequencing reads of the target sequence of the generated nucleic acid fragments. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original dsDNA molecule and, therefore, are no true duplicate reads. The ligation of random sequences onto dsDNA molecules, accompanied by the methods of the invention, allow for the identification of true duplicate reads versus apparent or perceived duplicate reads.

[0067] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acid fragments generated from a genomic DNA (gDNA) molecule that have the same start site, which can be determined from the sequencing reads of the target sequence of the nucleic acid fragments is determined. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original gDNA molecule and, therefore, are not true duplicate reads. In another embodiment, an identifier site is inserted at the adapter insert junction. The sequence of the identifier site is carried through the library amplification step. The identifier site is the first sequence read during the forward read. As the identifier sequence is not logically present adjacent to naturally occurring sequence, this uniquely identifies the DNA fragment. Therefore, by ligating random sequences onto the original gDNA, the methods of the invention identify true duplicate reads.

[0068] In some embodiments, a duplicate sequencing read is detected and analyzed. Duplicate reads can be filtered using 'samtools rmdup', wherein reads with identical external coordinates

are removed, only one read with highest mapping quality is retained. After filtering, a filtered set of deduplicated reads can be used in any downstream analysis. Conversely, this filtering step can be skipped, and downstream analysis can be done using the unfiltered reads, including duplicates.

[0069] In some embodiments, sequencing reads are generated from a number of samples. In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from the samples, in which the nucleic acid fragments from each sample has the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before sequencing reads of the target sequence and/or identifier site are analyzed. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before duplicate sequencing reads are analyzed and/or removed.

[0070] In some embodiments, a method disclosed herein identifies or detects one or more true duplicate(s) with increased accuracy as compared to other methods. For example, in some embodiments, a method disclosed herein identify true duplicates (in contrast to identifying apparent or perceived duplicates) with increased accuracy as compared to other methods. The increased resolution and/or accuracy in identifying one or more true duplicate(s) can provide a considerable contribution to the state of the art in more accurately identifying true duplicates. In some embodiments, a method disclosed herein identifies or detects a true duplicate with

increased efficiency as compared to other methods, such as paired end sequencing. The increase in accuracy, resolution, and/or efficiency in detecting duplicate reads (e.g., true duplicate(s)) can increase confidence in sequencing results, such as for expression and CNV analysis.

Kits

[0071] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for amplification.

[0072] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[0073] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit may be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[0074] A kit may include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

[0075] In some embodiments, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, a kit comprises a composition of the invention, in one or more containers. In some embodiments, the

invention provides kits comprising adapters, primers, and/or other oligonucleotides described herein. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) a forward adapter (e) one or more oligonucleotides comprising reverse adaptor sequence and (f) one or more buffers suitable for one or more of the elements contained in said kit. The adapters, primers, other oligonucleotides, and reagents can be, without limitation, any of those described above. Elements of the kit can further be provided, without limitation, in any of the amounts and/or combinations (such as in the same kit or same container) described above. The kits may further comprise additional agents, such as those described above, for use according to the methods of the invention. For example, the kit can comprise a first forward adaptor that is a partial duplex adaptor as described herein, a second forward adapter, and a nucleic acid modifying enzyme specific for a restriction and/or cleavage site present in the first forward adaptor. The kit elements can be provided in any suitable container, including but not limited to test tubes, vials, flasks, bottles, ampules, syringes, or the like. The agents can be provided in a form that may be directly used in the methods of the invention, or in a form that requires preparation prior to use, such as in the reconstitution of lyophilized agents. Agents may be provided in aliquots for single-use or as stocks from which multiple uses, such as in a number of reaction, may be obtained.

[0076] In some embodiments, the kit comprises a plurality of adaptor oligonucleotides, wherein each of the adaptor oligonucleotides comprises at least one of a plurality of identifier site sequences, wherein each identifier site sequence of the plurality of identifier site sequences differs from every other identifier site sequence in said plurality of identifier site sequences at at least three nucleotide positions, and instructions for using the same. Adapters comprising

different identifier site sequences can be supplied individually or in combination with one or more additional adapters having a different identifier site sequence. In some embodiments, the kit can comprises a plurality of adapter oligonucleotides.

EXAMPLES

Example 1: Identification of Duplicate Sequencing Reads with NuGEN Ovation Target Enrichment Library System

[0077] Sample Description: 100ng of DNA from a human HapMap sample (NA19238) was fragmented to approximately 500 base pair in length by sonication with a Covaris system (Covaris, Inc., Woburn, MA). The resulting DNA was treated with end repair enzyme mix NuGEN R01280 and R01439 (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendation to produce blunt ended DNA fragments.

[0078] Library Generation, Enrichment, and Identifier Site Incorporation: An oligonucleotide with segments, from 5' to 3' of the top strand, 1) an Illumina indexing read priming site such as AGAGCACACGTCTGAACTCCAGTCAC (SEQ ID NO:2), 2) an indexing site, 3) an identifier site having a random 6 base sequence and, 4) a sequence compatible with an Illumina forward sequencing priming site such as TCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO:3) , was annealed to a second oligonucleotide to form a partially double stranded DNA adapter. Five uM of these adapters were ligated onto the end-repaired DNA using Ligase and Ligase reaction buffer from NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendations. Following 30 minutes of incubation at 25°C, the reaction mixture was diluted with water, 0.8X volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated

DNA fragments eluted according to manufacturer's recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95°C then slowly cooling the mixture from 80°C to 60°C by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, MA) according to the manufacturer's protocols. Following extension, the DNA fragments were collected on Agencourt magnetic beads, washed and eluted according to manufacturer's recommendations. These libraries were enriched by 30 cycles of PCR using NuGEN library enrichment primers (Ovation Target Enrichment Library System, NuGEN Technologies, Inc., San Carlos, CA) that also contain the Illumina flow cell sequences (Illumina Inc., San Diego, CA) according to supplier's recommendation.

[0079] The resulting libraries were quantitated by qPCR using a kit provided by KAPA, diluted to 2nM and applied to an Illumina MiSeq DNA Sequencer (Illumina Inc., San Diego, CA). The following series was run: 36 base first read, 14 base second read, and 24 base third read.

[0080] *Data Analysis:* The sequencer output was processed in accordance with manufacturer's recommendation. In order to analyze the data, the indexing read was split into two files. The first file contained the first 8 bases of the indexing reads and is utilized as the library index file for standard library parsing. The other file contains only the random bases and is set aside for further sequence parsing.

[0081] Following our data analysis pipeline of sequence alignment with bowtie aligner (Langmead B. et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009, 10:R2.), duplicate reads were identified by their genomic start positions. At this point, sequencing reads that started at the same genomic position were checked against the random bases file to see if they have the same or different set of random

bases that were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation Target Enrichment targeting probe was used to generate the sequence fragment in question. These sequences, therefore, did not provide unique information about the starting genomic DNA and are considered as one sequencing read for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. **Figure 5** provides the analysis results demonstrating the identification of the duplicate reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

[0082] If the sequences of two libraries are identical, their duplication status is unknown since this could occur by chance in any library. If an identifier sequence is used in combination with the library reads, the status can be determined (duplicates if identical, distinct if different). With the SPET system, one end is common so the probability of two libraries having identical ends increases. These would appear to be duplicate sequences and their true status can be determined by looking at an identifier sequence.

[0083] Over the sampling of all randomly selected reads, the use of identifier sites provided an increased resolution on the presence of true duplicates. When evaluating two million random reads, the apparent duplicates were found to comprise 39% of all reads. However, the true duplicates, identified through the use of an identifier site, were found to comprise only 26% of all reads. The methods employing the use of an identifier site were found to considerably increase the resolution of the true number of duplicates within the pool of reads.

Example 2: Removal of Duplicate Sequencing Reads with 8 Base Identifier Site

[0084] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0085] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites. This identifier site is designed as eight random bases introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain the 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rmdup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 3: Removal of Duplicate Sequencing Reads with Mixture of Random 1-8 Base Identifier Sites

[0086] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the

ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0087] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0088] One to eight random bases are introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain between 1 and 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rndup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 4: Removal of Duplicate Sequencing Reads with Mixture of 96 Defined 6-Base Identifier Sites

[0089] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0090] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0091] A mixture of 96 defined six-bases sequences are introduced onto the end of the adaptors. Thus, each six-base sequence is an identifier site. Sequence reads from libraries made with such adaptors contain one of the 96 six-base identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rndup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 5: Identification of Duplicate Sequencing Reads in Determining mRNA Expression Levels

[0092] Sample Description: Total RNA is extracted from tumor and normal adjacent tissue for the purpose of finding differences in expression levels of transcripts between the two sample types. 100ng of each sample is converted into cDNA using the USP primers, reaction buffer, and Reverse Transcriptase provided in NuGEN's Encore Complete Library System (NuGEN Technologies, Inc., San Carlos, CA) according to the supplier's recommendations. This is followed by second strand synthesis, again using materials provided in the kit according to recommendations. Double stranded cDNA was prepared using the SuperScript® Double-Stranded cDNA Synthesis Kit from Life Technologies (Carlsbad, CA) according the manufacturer's instructions. DNA was sheared with a Covaris S-series device (Covaris, Inc., Woburn, MA) using the 200 bp sonication protocol provided with the instrument (10% duty cycle, 200 cycles/burst, 5 intensity, 180 seconds). DNA was treated with 1.5 uL 10X Blunting Buffer, 0.5 uL Blunting Enzyme (New England Biolabs, Inc., Ipswich, MA; p/n E1201) and 1.2 uL of 2.5mM of each dNTP mix in a total volume of 15 uL for 30 minutes at 25°C followed by 10 minutes at 70°C.

[0093] *Library Generation*: The DNA fragments were then subjected to end repair using end repair buffers and enzymes provided in NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA).

[0094] Forward adaptor 5'

[0095] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTNNNNNNNN (SEQ ID NO:4),

[0096] Reverse adaptor 1) 5'

[0097] CAAGCAGAAGACGGCATACGAGATTCCCTTGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCTNNNNNNNN (SEQ ID NO:5),

[0098] Reverse adaptor 2) 5'

[0099] CAAGCAGAAGACGGCATACGAGATTGAAGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCNNNNNNNN (SEQ ID NO:6), and

[00100] Common partner 5' NNNNNNNNAGATCGGAAGAGC (SEQ ID NO:7) were all ordered from IDT (Integrated DNA Technologies, Coralville, IA). The reverse adaptors each contain a unique identifier (underlined) that enables libraries made with these adapters to be distinguished. In this case N represents an equimolar mixture of A, C, G, and T. A mixture of 5uM forward, 5uM reverse, and 10uM common in 10mM MgCl2, 50mM Tris pH 8 was heated to 95C for 5 minutes, then cooled to 20C. Adaptor ligation was performed by addition of 4.5 uL water, 3 uL Adaptor mix (prepared above), 6 uL 5X NEBNext Quick Ligation Reaction Buffer and 1.5 uL Quick T4 DNA Ligase (New England Biolabs, Inc., Ipswich, MA; p/n E6056), followed by incubation for 30 minutes at 25°C followed by 10 minutes at 70°C. Ligation products were purified by adding 70uL water and 80 uL of Ampure XP beads (Agencourt Genomics), washing twice with 70% ethanol and eluting with 20 uL of 10mM Tris pH 8.0.

Library products were amplified in a 50uL PCR containing 0.5uM each of primer (5' AATGATACGGCGACCACCGA (SEQ ID NO:8) , and 5' CAAGCAGAAGACGGCATACGA (SEQ ID NO:9), 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM MgC12, 0.2mM each dNTP, and 1 unit Taq polymerase. The reaction was cycled 15 times under the conditions 95C for 15 seconds, 60C for 1 minute. PCR products were purified with 1 volume of Ampure XP beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) as described above. The library was analyzed by HS DNA Bioanalyzer (Agilent Technologies, Santa Clara, CA) and quantitated with the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA; p/n KK4835) according to the supplied instructions. The resulting libraries are combined and are compatible with standard TruSeq single end or paired Illumina sequencing protocols for GAIIx, MiSeq, or Hi Seq sequencing instruments (Illumina Inc., San Diego, CA). The following series is run; 50 base first read, 6 base second read. A third read is not required for counting purposes or duplication analysis.

[00101] Data Analysis: The sequencer output was processed in accordance with manufacturer's recommendation. The 6 bases of the index read is used for standard library parsing, separating the data files from the two sample types. The first 50 bases of the target sequence read are compared to each other. Any read that has identical sequence to any other read is identified as a duplicate and removed from the population, thus, only a single copy is retained within the file. Once the duplicate reads have been removed, 8 bases are trimmed from each read. The trimmed reads are aligned to a reference genome. Differential expression is then determined by comparing FPKM (fragments per kilobase per million reads) values between libraries utilizing scripts such as cufflinks or cuffdiff (Trapnell et al. 2010, Nature Biotechnology, 28, 511-515; Trapnell et al. 2013, Nature Biotechnology, 31, 46-53).

Example 6: Sequencing of the Identifier Site With the Target Sequence in Paired End Reads

[00102] The Ovation Library System for Low Complexity Samples (NuGEN Technologies, Inc., San Carlos, CA) was used to generate four libraries, each from a single amplicon, following manufacturer's protocol. Purified libraries were mixed and sequenced as a multiplex on the Illumina MiSeq (Illumina Inc., San Diego, CA) to produce 125nt forward, 8nt index 1, 8nt index 2, and 25nt reverse reads. Because all amplicon reads start and end at the same sequence coordinates, the traditional method of marking library PCR duplicates (marking reads that start and end at the same genomic coordinates as duplicates) cannot be used. Instead, the 0-8nt of random sequence contained in the adaptors ligated to the amplicon were treated as an identifier sequence and used to mark duplicates. Any paired end reads that shared the same length and sequence of these random bases with any other paired end read was called a duplicate. The table below shows the results of this duplicate marking.

Table 1.

Library	Reads	Duplicate Reads	Reads from independent molecules
1	454249	376797	77452
2	439367	364001	75366
3	317760	253057	64703
4	476572	398380	78192

[00103] Table 1 demonstrates the accuracy of the method used to differentiate between duplicate reads and reads from truly independent molecules. The population of reads from independent molecules, depicted in the final column of Table 1, represent sequences from independent amplicon molecules used in generating the libraries.

Example 7: Sequencing of the Identifier Site With the Target Sequence in Reduced Representation Bisulfite (RRBS) Libraries

[00104] Reduced representation bisulfite (RRBS) libraries of the human genome are generated through the complete restriction enzyme digestion of 100ng input sample, followed by selection for short fragments. The resulting pool of fragments are ligated to adaptor sequences comprising an indexing site and an identifier site. The identifier sites comprise either 6 or 8 random nucleotides. The sequences are then sequenced to identify the identifier sites, thus revealing the true number of duplicates in the pool. In the absence of identifier sites, the number of apparent or perceived duplicates are greater than the number of true duplicates. The inclusion of an identifier site results in the identification of the number of true duplicates, as compared to the larger number of apparent or perceived duplicates.

[00105] Unless defined otherwise, all technical and scientific terms herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials, similar or equivalent to those described herein, can be used in the practice or testing of the present invention, the preferred methods and materials are described herein. All publications, patents, and patent publications cited are incorporated by reference herein in their entirety for all purposes.

[00106] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

[00107] All references, articles, publications, patents, patent publications, and patent applications cited herein are incorporated by reference in their entireties for all purposes. However, mention of any reference, article, publication, patent, patent publication, and

patent application cited herein is not, and should not be taken as an acknowledgment or any form of suggestion that they constitute valid prior art or form part of the common general knowledge in any country in the world.

[00108] While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth and as follows in the scope of the appended claims.

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

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 14/540,917

filed on 11/13/2014.

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 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.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Ben Schroeder Date (Optional) : _____

Signature: _____

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. 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 1 minute 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.

American LegalNet, Inc.
www.FormsWorkFlow.com

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

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 14/540,917

filed on 11/13/2014.

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 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.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Douglas Amorese Date (Optional) :

Signature:

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. 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 1 minute 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.

American LegalNet, Inc.
www.FormsWorkFlow.com

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

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 14/540,917

filed on 11/13/2014.

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 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.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Jonathan Scolnick Date (Optional) : _____

Signature: _____

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. 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 1 minute 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.

American LegalNet, Inc.
www.FormsWorkFlow.com

Amendments to the Specification

Please amend the "CROSS-REFERENCE TO RELATED APPLICATIONS" section at the top of page 1 as follows:

[0001] This application is a continuation of U.S. Patent Application Serial No. 15/406,002, filed January 13, 2017, which application is a Division of U.S. Application 14/540,917, filed November 13, 2014, now issued as U.S. Pat. 9,546,399, each of which claims the benefit of U.S. Provisional Application Serial No. 61/903,826, filed November 13, 2013, each of which is herein incorporated by reference in its entirety.

Electronic Acknowledgement Receipt

EFS ID:	38298349
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	15-JAN-2020
Filing Date:	
Time Stamp:	15:46:38
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1820
RAM confirmation Number	E20201EF46528158
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		NUGN-001-03US-Prlm-Amdt-011520.pdf	127886 5f422297bd6592b4645bd5e41ce2d054210b8f30	yes	3

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Applicant Arguments/Remarks Made in an Amendment	3	3
	Specification	2	2
	Preliminary Amendment	1	1

Warnings:

Information:

2		NUGN-001-03US-Application.pdf	2249576 b749306e55b6aa95b4e00bea98196a75895c405e	yes	50

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Drawings-only black and white line drawings	45	50
	Abstract	44	44
	Claims	41	43
	Specification	1	40

Warnings:

Information:

3	Oath or Declaration filed	NUGN-001-03US-Declarations.pdf	339534 dccecf84056f6bd25e501e9dac8593813d9374f6	no	3

Warnings:

Information:					
4	Application Data Sheet	NUGN-001-03US-ADS.pdf	1256192 7ef873e3229e3d0d968af06774867f17881 56ee	no	9
Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	36776 856f597f700426809d76766eee87dd5f5a8a e3c8	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	4009964	

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

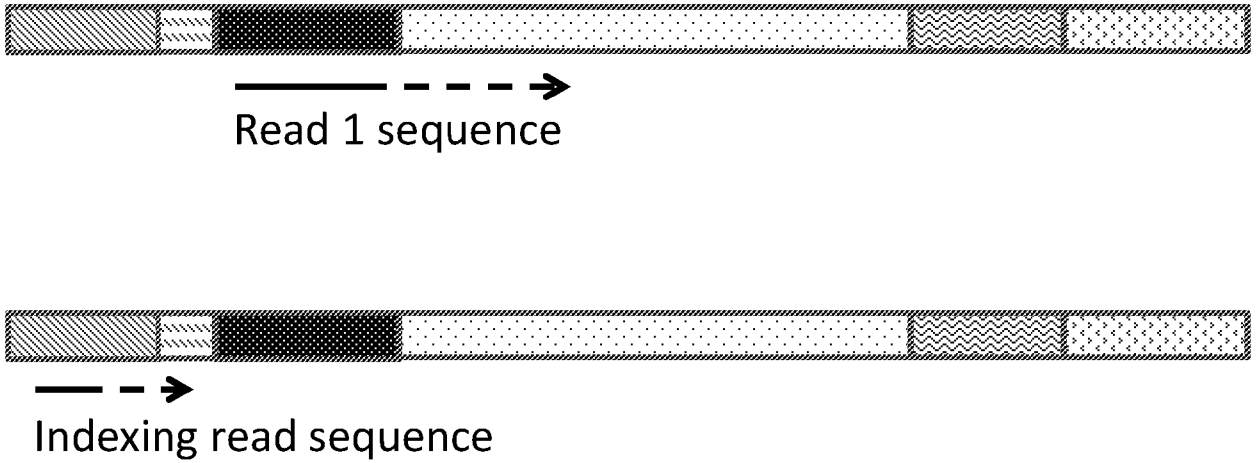
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will 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.

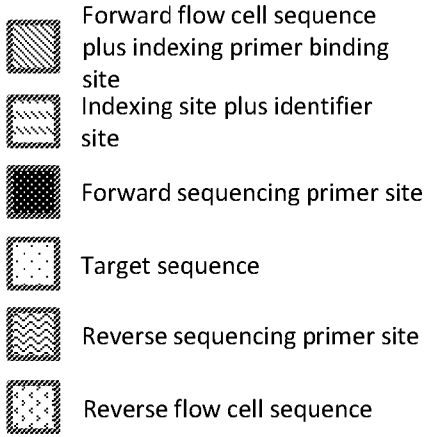
ABSTRACT

The present invention provides methods, compositions and kits for detecting duplicate sequencing reads. In some embodiments, the duplicate sequencing reads are removed.

Standard DNA Sequencing With Duplication Counter

- Sequence Read 1 (forward read)





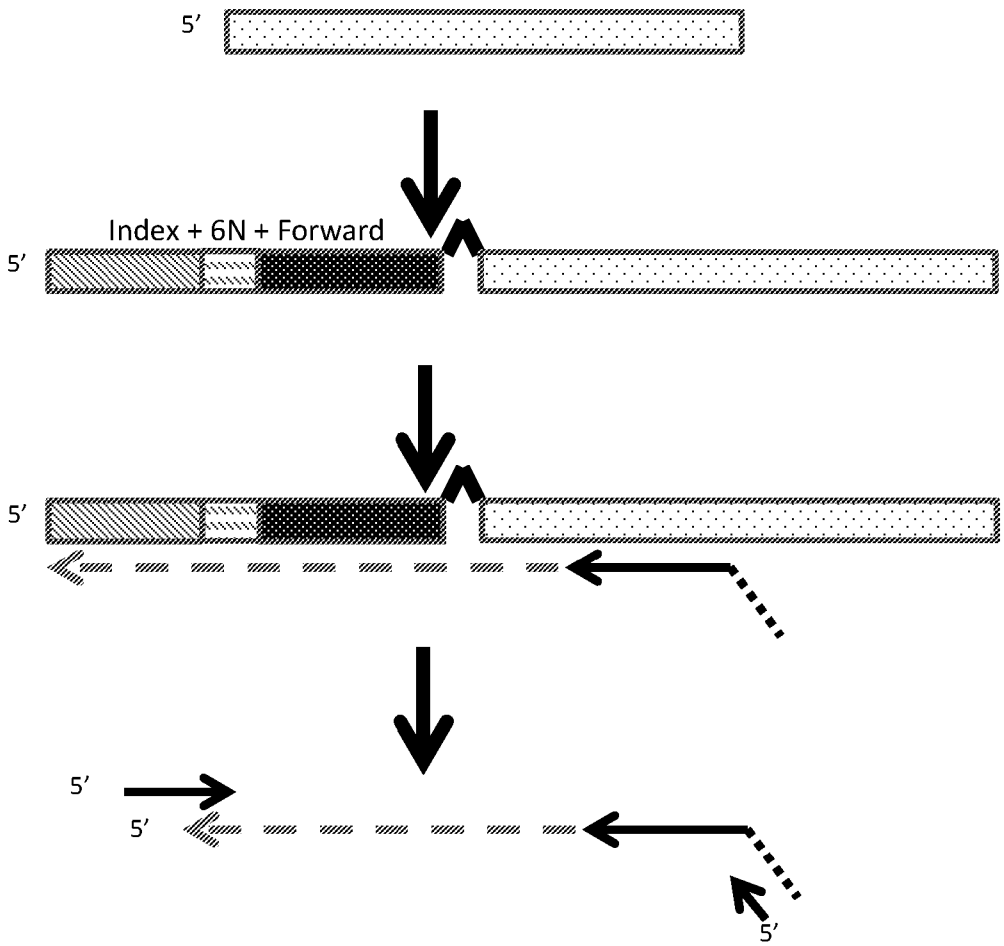
The indexing read results in sequencing of both the index site and the identifier site

- Extension of the indexing read beyond the indexing site and into the nucleotides of the identifier site provides data on which nucleic acid molecule is being sequenced.

Figure 1

IDT Exhibit 1019.0065

Single Primer Extension (SPET) Reaction



gDNA is sheared

Adapters are ligated on to end-repaired DNA. The adapters contain at least a partial indexing sequence priming site, an indexing site, an identifier site, and the forward sequencing priming site. Each gDNA is now labeled with a unique identifier sequence
Targeting oligos are hybridized to gDNA and extended through the adapter containing an identifier site (6N) sequence

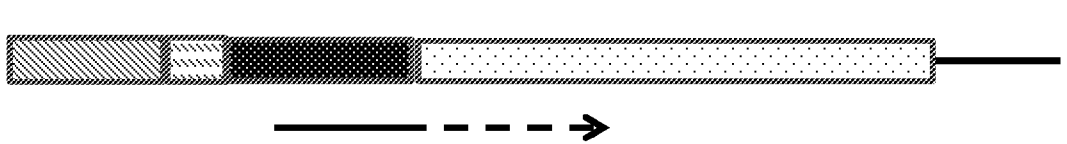
Probes that have extended through an adapter are PCR amplified for sequencing

Figure 2A

IDT Exhibit 1019.0066

SPET Sequencing

- Sequence from random end toward the probe



PCR adds sequence specific primer binding sites and Read 1 of sequencing is performed

Read 1 sequence



Indexing Read performed sequences both the indexing site as well as the identifier site (N6)

Indexing read sequence

- Extending the indexing read past the 6 base indexing site sequence and into the random Ns of the identifier site provides data on which gDNA molecule the extension reaction occurred upon.

Figure 2B

IDT Exhibit 1019.0067

Example Adaptor

An adaptor sequence with necessary priming sites, the index site, and the identifier site

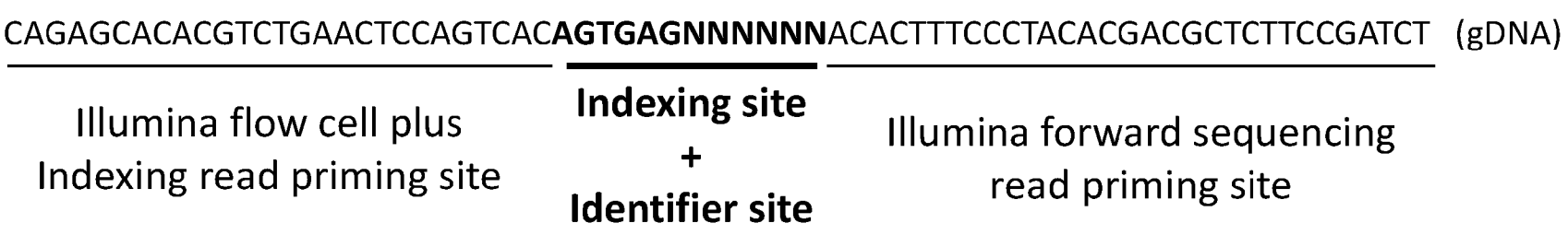
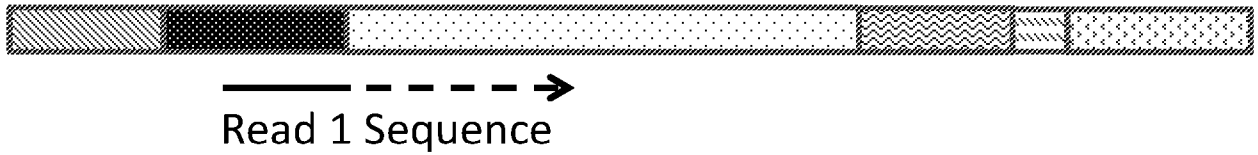


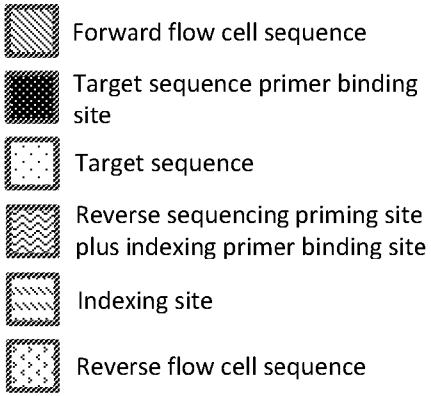
Figure 3

IDT Exhibit 1019.0068

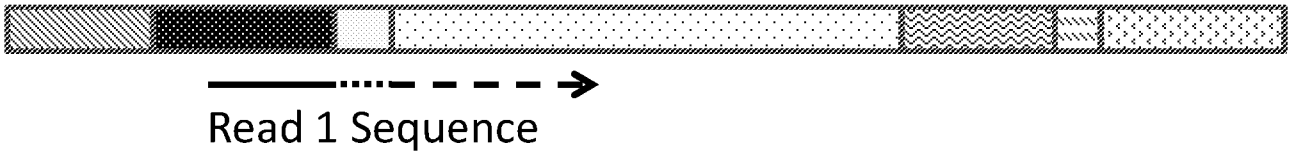
DNA Sequencing With Duplicate Recognition at the insert junction

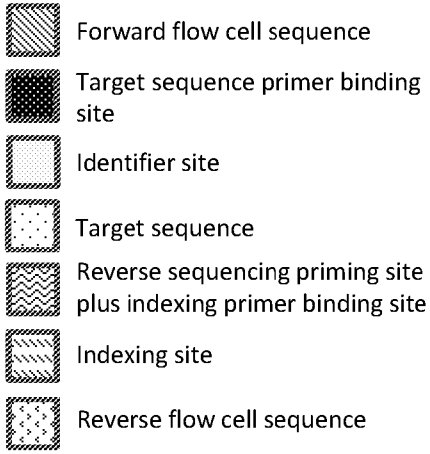
- Traditional library, sequence read 1 (forward read)





- Duplicate recognition library, sequence read 1 (forward read)





Read 1 initially records the approximately 8 random bases of the identifier site appended to the library sequences, followed by the target sequence

Figure 4

IDT Exhibit 1019.0069

Resolution of apparent duplicate reads by identifier sequence ligation

Libraries were generated from 100ng genomic DNA Covaris fragmented to ~500bp in length. These were enriched with the 344 gene cancer panel, ~12k probes covering 969kb of genomic space and sequenced on the ILMN Miseq. Different numbers of reads were randomly selected for this analysis

Total reads randomly selected	Aligned read	Fold Target Coverage	Total duplicates after identifier analysis (true rate)	Percentage duplicates after identifier analysis (true rate)	Total apparent duplicates with no identifier	Percentage apparent duplicates with no identifier
100K	97,919	2X	1854	1.89	3256	3.33
500K	488,423	7X	39731	8.13	68741	14.07
1M	976,662	13X	145589	14.91	237882	24.36
2M	1,954,673	21X	514848	26.34	764825	39.13
10M	9,771,462	43X	6493577	66.45	7515496	76.91

"Apparent" Duplicates – individual targets within libraries consist of one random end (result of fragmentation) and one common end (initiated with the sequence specific probe). Apparent duplicates are derived from sequencing libraries and using conventional measures of duplicate reads ie, reads were mapped using bowtie and all reads with the same start and end genomic coordinates were counted as duplicates.

"True" duplicates - An identifier sequence is introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates. The analysis method is described in the specification.

Figure 5

IDT Exhibit 1019.0070

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
16/743,604	01/15/2020		1820	NUGN-001/03US 34079/516	21	2

CONFIRMATION NO. 3249

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

FILING RECEIPT

OC000000114795244

Date Mailed: 02/18/2020

Receipt is acknowledged of this non-provisional utility 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 FIRST INVENTOR, 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 corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction 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 provided that the request is grantable.

Inventor(s)

Douglas A. Amorese, Los Altos, CA;
Jonathan Scolnick, San Francisco, CA;
Benjamin G. Schroeder, San Mateo, CA;

Applicant(s)

NUGEN TECHNOLOGIES, INC., Redwood City, CA;

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 15/406,002 01/13/2017 PAT 10570448
which is a DIV of 14/540,917 11/13/2014 PAT 9546399
which claims benefit of 61/903,826 11/13/2013

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

IDT Exhibit 1019.0071

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: 02/14/2020
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/743,604**
Projected Publication Date: To Be Determined - pending completion of Sequence Disclosures
Non-Publication Request: No
Early Publication Request: No
Title

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

Preliminary Class

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

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

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

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

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

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

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

IDT Exhibit 1019.0072

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: 16/743,604

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			N/A	300
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A			N/A	660
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	760
TOTAL CLAIMS (37 CFR 1.16(i))	21 minus 20 =	* 1			OR	x 100 =	100
INDEPENDENT CLAIMS (37 CFR 1.16(h))	2 minus 3 =	*				x 460 =	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			TOTAL	1820

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	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
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
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516

CONFIRMATION NO. 3249

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

FORMALITIES LETTER

OC000000114755245

Date Mailed: 02/18/2020

NOTICE TO COMPLY WITH REQUIREMENTS FOR PATENT APPLICATIONS CONTAINING NUCLEOTIDE AND/OR AMINO ACID SEQUENCE DISCLOSURES

Filing Date Granted

Applicant is given **TWO MONTHS FROM THE DATE OF THIS NOTICE** within which to comply with the sequence rules, 37 CFR §§ 1.821-1.825. Failure to comply with these requirements will result in ABANDONMENT of the application under 37 CFR §1.821(g). Extension of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR §1.136.

Items 1) and 2) provide general guidance related to requirements for sequence disclosures.

1) 37 CFR 1.821(c) requires that patent applications which contain disclosures of nucleotide and/or amino acid sequences that fall within the definitions of 37 CFR 1.821(a) must contain a "Sequence Listing", *as a separate part of the disclosure*, which presents the nucleotide and/or amino acid sequences and associated information using the symbols and format in accordance with the requirements of 37 CFR 1.821-1.825. This "Sequence Listing" part of the disclosure may be submitted:
   a) via EFS-Web (see Section I.1 of the Legal Framework for EFS-Web (https://www.uspto.gov/patents-application-process/filing-online/legal-framework-efs-web), hereinafter "Legal Framework") as an ASCII text file, together with an incorporation-by-reference of the material in the ASCII text file in a separate paragraph of the specification identifying:
      i) the name of the ASCII text file;
      ii) the date of creation; and
      iii) the size of the ASCII text file in bytes;
   b) on compact disc(s) in duplicate according to 37 CFR 1.52(e)(1)(ii) and (4), labeled according to 37 CFR 1.52(e)(6), together with a statement that the duplicate compact discs are identical and an incorporation-by-reference of the material in the ASCII text file according to 37 CFR 1.52(e)(5) in a separate paragraph of the specification identifying:
      i) the name of the ASCII text file;
      ii) the date of creation; and
      iii) the size of the ASCII text file in bytes;
   c) via EFS-Web as a PDF (not recommended); or
   d) on paper.
2) 37 CFR 1.821(e) requires that a *copy* of the "Sequence Listing" must also be submitted in computer readable form (CRF) in accordance with the requirements of 37 CFR 1.824.
   a) If a "Sequence Listing" ASCII text file submitted via EFS-Web on the application filing date complies with the requirements of 37 CFR 1.824(a)(2)-(6) and (b), and applicant has not filed a "Sequence Listing" as

IDT Exhibit 1019.0075

a PDF, the text file will serve as both the "Sequence Listing" required by 37 CFR 1.821(c) and the CRF required by 37 CFR 1.821(e), and the statement of identity under the "Legal Framework" is not required.

b) If the "Sequence Listing" required by 37 CFR 1.821(c) is filed via EFS-Web as a PDF, then the "Legal Framework" requires submission of a statement that the "Sequence Listing" content of the PDF copy and the ASCII text file copy submitted via EFS-Web are identical.

c) If the "Sequence Listing" required by 37 CFR 1.821(c) is filed on paper or compact disc, then 37 CFR 1.821(f) requires submission of a statement that the "Sequence Listing" content of the paper or compact disc copy and the CRF are identical.

Specific deficiencies and the required response to this notice are as follows:

- This application fails to comply with the requirements of 37 CFR 1.821-1.825 because it does not contain a "Sequence Listing" as a separate part of the disclosure or a CRF. Applicant must provide:
  - A "Sequence Listing" part of the disclosure, as described above in item 1); **together with**
  - **An amendment specifically directing its entry into the application;**
  - A statement that the "Sequence Listing" includes no new matter; and
  - A statement that indicates support for the amendment in the application, as filed, as required by 37 CFR 1.825 and the "Legal Framework".
- If the "Sequence Listing" part of the disclosure is submitted according to item 1) a) or b) above, Applicant must also provide:
  - A substitute specification in compliance with 37 CFR 1.121(b)(3) and 1.125 inserting the required incorporation-by-reference paragraph, consisting of:
    - A copy of the previously-submitted specification, with deletions shown with strikethrough or brackets and insertions shown with underlining (marked-up version);
    - A copy of the amended specification without markings (clean version); and
    - A statement that the substitute specification contains no new matter.
- If the "Sequence Listing" part of the disclosure is submitted according to item 1) b), c), or d) above, Applicant must also provide:
  - A CRF; and
  - A statement according to item 2) b) or c) above.
- If applicant desires the CRF in this application to be identical with the CRF 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 (Form PTO/SB/93 may be used for this purpose).

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**
- **PatentIn software is available at http://www.uspto.gov/patents/resources/tools/checker/patentinrel.jsp .**

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.

IDT Exhibit 1019.0076

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**.

/juwimana/

IDT Exhibit 1019.0077

==========

Sequence Listing was accepted.

If you need help call the Patent Electronic Business Center at (866) 217-9197 (toll free).

Reviewer: Anjum, Durreshwar

Timestamp: [year=2020; month=4; day=17; hr=13; min=32; sec=52; ms=219;]

==========

Validated By CRFValidator v 1.0.5

Application No: 16743604 Version No: 1.0

Input Set:

Output Set:

Started: 2020-04-17 13:05:55.234
Finished: 2020-04-17 13:05:56.064
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 830 ms
Total Warnings: 47
Total Errors: 0
No. of SeqIDs Defined: 9
Actual SeqID Count: 9

Error code		Error Description
W	213	Artificial or Unknown found in <213> in SEQ ID (1)
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	213	Artificial or Unknown found in <213> in SEQ ID (2)
W	213	Artificial or Unknown found in <213> in SEQ ID (3)
W	213	Artificial or Unknown found in <213> in SEQ ID (4)
W	447	n or Xaa used, for: SEQID(4) on line number 85
W	447	n or Xaa used, for: SEQID(4) on line number 85
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	213	Artificial or Unknown found in <213> in SEQ ID (5)
W	447	n or Xaa used, for: SEQID(5) on line number 107

Input Set:

Output Set:

Started: 2020-04-17 13:05:55.234
Finished: 2020-04-17 13:05:56.064
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 830 ms
Total Warnings: 47
Total Errors: 0
No. of SeqIDs Defined: 9
Actual SeqID Count: 9

Error code	Error Description
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107 This error has occured more than 20 times, will not be displayed
W 213	Artificial or Unknown found in <213> in SEQ ID (6)
W 213	Artificial or Unknown found in <213> in SEQ ID (7)
W 213	Artificial or Unknown found in <213> in SEQ ID (8)
W 213	Artificial or Unknown found in <213> in SEQ ID (9)

SEQUENCE LISTING

<110> NuGEN Technologies, Inc.

<120> COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCE READ

<130> NUGN-001/03US 34079/516

<140> US 16/743,604
<141> 2020-01-15

<150> 15/406,002
<151> 2017-01-13

<150> 14/540,917
<151> 2014-11-13

<150> 61/903,826
<151> 2013-11-13

<160> 9

<170> PatentIn version 3.5

<210> 1
<211> 70
<212> DNA
<213> artificial

<220>
<223> Adaptor

<220>
<221> misc_feature
<222> (34)..(39)
<223> n is a, c, g, or t

<400> 1
cagagcacac gtctgaactc cagtcacagt gagnnnnnna cactttccct acacgacgct 60

cttccgatct 70

<210> 2
<211> 26
<212> DNA
<213> artificial

<220>
<223> Priming site

<400> 2
agagcacacg tctgaactcc agtcac 26

<210> 3
<211> 29
<212> DNA
<213> artificial

<220>
<223> forward priming site compliment

<400> 3
tctttcccta cacgacgctc ttccgatct 29

<210> 4
<211> 66
<212> DNA
<213> artificial

<220>
<223> Forward adaptor

<220>
<221> misc_feature
<222> (59)..(66)
<223> n is a, c, g, or t

<400> 4
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctnn 60

nnnnnn 66

<210> 5
<211> 72
<212> DNA
<213> artificial

<220>
<223> Reverse adaptor

<220>
<221> misc_feature
<222> (65)..(72)
<223> n is a, c, g, or t

<400> 5
caagcagaag acggcatacg agattccctt gtgactggag ttcagacgtg tgctcttccg 60

atctnnnnnn nn 72

<210> 6
<211> 71
<212> DNA
<213> artificial

<220>
<223> Reverse Adaptor

<220>
<221> misc_feature
<222> (64)..(71)
<223> n is a, c, g, or t

<400> 6
caagcagaag acggcatacg agattgaagg gtgactggag ttcagacgtg tgctcttccg 60

atcnnnnnnn n 71

<210> 7
<211> 21
<212> DNA
<213> artificial

<220>
<223> Common partner

<220>
<221> misc_feature
<222> (1)..(8)
<223> n is a, c, g, or t

<400> 7
nnnnnnnnag atcggaagag c 21

<210> 8
<211> 20
<212> DNA
<213> artificial

<220>
<223> Primer

<400> 8
aatgatacgg cgaccaccga 20

<210> 9
<211> 21
<212> DNA
<213> artificial

<220>
<223> Primer

<400> 9
caagcagaag acggcatacg a 21

DocCode – SEQ.TXT

SCORE Placeholder Sheet for IFW Content

Application Number: 16743604 Document Date: 04/17/2020

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

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

- Sequence Listing

At the time of document entry (noted above):

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

Form Revision Date: March 1, 2019

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPFICATE SEQUENCING READ

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. Patent Application Serial No. 15/406,002, filed January 13, 2017, which application is a Division of U.S. Application 14/540,917, filed November 13, 2014, now issued as U.S. Pat. 9,546,399, each of which claims the benefit of U.S. Provisional Application Serial No. 61/903,826, filed November 13, 2013, each of which is herein incorporated by reference in its entirety.

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0002] The contents of the text file submitted electronically herewith are incorporated herein by reference in their entirety: A computer readable format copy of the Sequence Listing (filename: ~~NUGN_001_01US_SeqList_ST25.txt~~ NUGN-001-03US_SeqList_ST25.txt, date recorded: ~~November 12, 2014~~ April 3, 2020, file size 3 kilobytes).

FIELD OF THE PRESENT INVENTION

[0003] The present invention relates generally to the field of high throughput sequencing reactions and the ability to discern artifacts arising through sequence duplications from nucleotide molecules that are unique molecules.

BACKGROUND OF THE PRESENT INVENTION

[0004] In RNA sequencing applications, accurate gene expression measurements may be hampered by PCR duplicate artifacts that occur during library amplification. When analyzing RNA sequencing data, when two or more identical sequences are found, it can be difficult to know if these represent unique cDNA molecules derived independently from different RNA molecules, or if they are PCR duplicates derived from a single RNA molecule. In genotyping by

sequencing, duplicate reads can be considered non-informative and may be collapsed down to a single read, thus reducing the number of sequencing reads used in final analysis. Generally, sequencing reads may be determined to be duplicates if both forward and reverse reads have identical starting positions, even though two independently generated molecules can have identical starting positions by random chance. Single primer extension based targeted re-sequencing suffers from an issue in that only one end of a sequencing read is randomly generated, while the other (reverse read) end is generated by a specific probe. This may make it difficult to determine if two reads are duplicates because they have been duplicated by PCR or because by chance they happened to start at the same position.

[0005] In expression analysis studies there may be limited value in doing paired end sequencing since the goal of the experiment is to determine amounts of transcript present as opposed to studying exon usage. In these studies, paired end sequencing adds costs while the only value is in helping distinguish PCR duplicates. The probability of two reads starting in the same position on only one end is higher than the probability of two reads having the same starting position on two ends (forward and reverse read). There is a need for improved methods that allow for low-cost, high throughput sequencing of regions of interest, genotyping or simple detection of RNA transcripts without inherent instrument inefficiencies that drive up sequencing costs due to the generation of unusable or non-desired data reads. The invention described herein fulfills this need. Here, we describe an adaptor approach that allows for the identification of true PCR duplicates and their removal.

[0006] The methods of the present invention provide novel methods for identifying true duplicate reads during sequencing, such as to improve data analysis of sequencing data, and other related advantages.

SUMMARY OF THE PRESENT INVENTION

[0007] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The detection and/or removal of duplicate sequencing reads presented herein is a novel approach to increasing the efficacy of evaluating data generated from high throughput sequence reactions, including complex multiplex sequence reactions.

[0008] Accordingly, the present invention provides a method of detecting a duplicate sequencing read from a population of sample sequencing reads, the method comprising ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site. The ligated adaptor-nucleic acid fragment products can be amplified, thus generating a population of sequencing reads from the amplified adaptor-nucleic acid ligation products. The sequencing reads with a duplicate identifier site and target sequence can then be detected from the population of sequencing reads. The methods can further include the removal of the sequencing reads with the duplicate identifier site and target sequence from the population of sequence reads.

[0009] In some embodiments, the identifier site is sequenced with the indexing site or the target sequence. In further embodiments, the identifier site is sequenced separately from the indexing site or the target sequence.

[0010] In some embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the identifier site; and the target sequence primer binding site. In further embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the target sequence primer binding site; and the identifier site.

[0011] In some embodiments, the plurality of nucleic acid fragments is generated from more than one sample. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, the sequencing reads are separated based on the indexing site. In yet other embodiments, the separation of sequencing reads is performed prior to detecting sequence reads with a duplicate identifier site and target sequence.

[0012] In some embodiments, the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments. In further embodiments, the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

[0013] In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In further embodiments, the indexing site is about 6 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In further embodiments, the identifier site is about 8 nucleotides in length.

[0014] In some embodiments, the indexing primer binding site is a universal indexing primer binding site; and in some embodiments, the target sequence primer binding site is a universal target sequence primer binding site.

[0015] The present invention also encompasses embodiments that include a kit comprising a plurality of adaptors, wherein each adaptor comprise an indexing primer binding site; an indexing site, and identifier site, and a target sequencing primer binding site.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] A better understanding of the novel features of the invention and advantages of the present invention will be obtained by reference to the following description that sets forth

illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0017] Figure 1 depicts a schematic of generating sequencing reads of a library including where an indexing primer and a target sequence primer anneal.

[0018] Figure 2A depicts a mechanism of the Single Primer Enrichment Technology (SPET) and how an identifier site is carried over into the final library and how sequencing through the indexing site into the identifier site provides data on which nucleic acid molecule is being identified.

[0019] Figure 2B is a continuation of Figure 2A.

[0020] Figure 3 offers a detailed view of the sequence of an example adaptor, among many envisioned embodiments, and the position of the indexing and identifier sites (SEQ ID NO: 1). "N" refers to any nucleic acid.

[0021] Figure 4 depicts a schematic of two separate sequence libraries, indicating where the indexing primers and the target primers anneal in a traditional library as compared to a library using an identifier site.

[0022] Figure 5 depicts a data table demonstrating the accuracy of resolving true duplicates versus apparent or perceived duplicates using an identifier site in the adaptors.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0023] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The present invention encompasses methods of detecting sequences duplicated in sequencing applications, and further removal of the duplicated sequence reads. The present invention further encompasses kits

comprising components that would allow for customized applications of the method of detecting and removing duplicated sequence reads in high throughput sequencing reactions. The compositions and methods can be used with various applications for genetic sample analysis, such as RNA sequence analysis, copy number variation analysis, methylation sequencing analysis, genotyping and whole genome amplification.

[0024] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the present invention. On the contrary, the present disclosure is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the present invention.

[0025] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg 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.

[0026] In some embodiments, the methods disclosed herein is for detecting from a population of sequencing reads a sequencing read, such as a duplicate sequencing read with a duplicate identifier site and target sequence. A duplicate sequencing read can be a sequencing read with

the same identifier site and target sequence as another sequencing read in the population of sequencing reads.

Adaptor

[0027] The present invention provides compositions of adaptors and methods comprising use of an adaptor. An adapter refers to an oligonucleotide sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. The target polynucleotide molecules may be fragmented or not prior to the addition of adaptors. In some embodiments, a method disclosed herein comprises ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples.

[0028] Various adaptor designs are envisioned which are suitable for generation of amplification-ready products of target sequence regions/strands of interest. For example, the two strands of the adaptor may be self-complementary, non-complementary or partially complementary. In some embodiments, the adaptor can comprise an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site.

[0029] An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site. An indexing site is a nucleic acid sequence that acts as an index for multiple polynucleotide samples, thus allowing for the samples to be pooled together into a single sequencing run, which is known as multiplexing. In some embodiments, the indexing site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In some embodiments, the indexing site is about 6 nucleotides in length.

[0030] An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads. In some embodiments, the identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In some embodiments, the identifier site is about 8 nucleotides in length. This identifier site can be designed as a set of sequences, or it can be semi-random, or it can be completely random. In addition, this identifier site can be a fixed length, or it can be a variable length. In some embodiments, the identifier sites in a plurality of adaptors are of a fixed length. For example, the identifier sites can all be eight random bases. In another embodiment, the identifier sites in a plurality of adaptors are of a variable length. For example, the identifier sites can range in size from 1 to 8 bases. In yet another embodiment, the identifier sites can be of a defined set of defined sequence. For example, the identifier site of a plurality of adaptors can be one of 96 defined six-base nucleotide sequences.

[0031] A target sequence primer binding site nucleotide sequence for binding a primer for a target sequence. The primer can be used to amplify the target sequence (e.g., a nucleic acid fragment from a sample). Accordingly, in some embodiments, the adaptor comprises an indexing primer binding site and a target sequence primer binding site.

[0032] A primer is a polynucleotide chain, typically less than 200 residues long, most typically between 15 and 100 nucleotides long, but can encompass longer polynucleotide chains. A primer targeting the primer binding sites are typically designed to hybridize to single-stranded nucleic acid strands. In some embodiments, the primers targeting the primer binding sites are designed to hybridize to single-stranded DNA targets. In the case where the sample comprises genomic DNA or other double-stranded DNA, the sample can be first denatured to render the target single stranded and enable hybridization of the primers to the desired sequence regions of

interest. In these embodiments, the methods and compositions described herein can allow for region-specific enrichment and amplification of sequence regions of interest. In some embodiments, the other double-stranded DNA can be double-stranded cDNA generated by first and second strand synthesis of one or more target RNAs.

[0033] In other embodiments, the primers targeting the primer binding sites are designed to hybridize to double-stranded nucleic acid targets, without denaturation of the double stranded nucleic acids. In other embodiments, the primers targeting the primer binding sites are designed to hybridize to a double-stranded DNA target, without denaturation of the dsDNA. In these embodiments, the primers targeting the selected sequence regions of interest are designed to form a triple helix (triplex) at the selected sequence regions of interest. The hybridization of the primers to the double-stranded DNA sequence regions of interest can be carried out without prior denaturation of the double stranded nucleic acid sample. In such embodiments, the methods and compositions described herein can allow for region-specific enrichment as well as strand-specific enrichment and amplification of sequence regions of interest. This method can be useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and analysis of multiplicity of sequence regions of interest in the native complex nucleic acid sample. The method can find use for studies and analyses carried out in situ, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as permit the analysis of complex genomic DNA without disruption of chromatin structures.

[0034] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as for amplification of a target sequence (e.g., in PCR). The oligonucleotide primer can be a synthetic polynucleotide

that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a sequence of the target polynucleotide. In some embodiments, the 3' region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a primer binding site.

[0035] In some embodiments, the primer binding site is a binding site for a universal primer. A universal primer is a primer that can be used for amplifying a number of different sequences. In some embodiments, a universal primer is used to amplify different libraries. In some embodiments, an indexing primer binding site is a binding site for a universal indexing primer (i.e., the indexing primer binding site is a universal indexing primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal indexing primer binding site. In some embodiments, the universal indexing primer can be used to amplify and/or sequence a number of different indexing sites.

[0036] In some embodiments, a target sequence primer binding site is a binding site for a universal target sequence primer (i.e., the target sequence primer binding site is a universal target sequence primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal target sequence primer binding site. In some embodiments, the universal target sequence primer can be used to amplify and/or sequence a number of different target sequences.

[0037] In some embodiments, the adaptor comprises an identifier site 3' to an indexing site. In some embodiments, the adaptor comprises an identifier site 5' to an indexing site. In some embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target sequence primer binding site. In other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target

sequence primer binding site. In yet other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, identifier site, indexing site, and target sequence primer binding site.

Samples

[0038] In some embodiments, an adaptor is ligated to a nucleic acid fragment (e.g., the 5' end of the nucleic acid fragment). The nucleic acid fragment can be from a plurality of nucleic acid fragments from one or more samples. The nucleic acid fragment can be RNA, DNA, or complex DNA, for example genomic DNA and PNA, in which case one might use a modified nucleic acid. The nucleic acid fragment may also be cDNA. The cDNA can be generated from RNA, e.g., mRNA.

[0039] The sample can be a biological sample. For example, the sample can be an animal, plant, bacterial, algal, or viral sample. In some embodiments, the sample is a human, rat, or mouse sample. The sample can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The sample can be cDNA made from a mixture of genomes of different species. In some embodiments, the sample can be from a synthetic source. The sample can be mitochondrial DNA. The sample can be cell-free DNA. The cell-free DNA can be obtained from sources such as a serum or a plasma sample. The sample can comprise one or more chromosomes. For example, if the sample is from a human, the sample can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. In some embodiments, the sample comprises a linear or circular genome. The sample can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The sample can be from more than one individual or organism. The sample

can be double-stranded or single-stranded. The sample can be part of chromatin. The sample can be associated with histones.

[0040] In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from more than one sample, such as 2, 3, 4, 5, or more samples. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site.

[0041] In some embodiments, a population of nucleic acid fragments generated from a sample is of one or more specific size range(s). In some embodiments, the fragments have an average length from about 10 to about 10,000 nucleotides. In some embodiments, the fragments have an average length from about 50 to about 2,000 nucleotides. In some embodiments, the fragments have an average length from about 100-2,500, 10-1,000, 10-800, 10-500, 50-500, 50-250, or 50-150 nucleotides. In some embodiments, the fragments have an average length less than 10,000 nucleotide, such as less than 5,000 nucleotides, less than 2,500 nucleotides, less than 2,500 nucleotides, less than 1,000 nucleotides, less than 500 nucleotides, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides.

[0042] In some embodiments, fragmentation of the nucleic acids can be achieved through methods known in the art. Fragmentation can be achieved through physical fragmentation methods and/or enzymatic fragmentation methods. Physical fragmentation methods can include

nebulization, sonication, and/or hydrodynamic shearing. In some embodiments, the fragmentation can be accomplished mechanically comprising subjecting the nucleic acids in the input sample to acoustic sonication. In some embodiments, the fragmentation comprises treating the nucleic acids in the input sample with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of nucleic acid or polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g., from New England Biolabs). For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some embodiments, fragmentation comprises treating the nucleic acids in the input sample with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel.

[0043] In some embodiments, fragmentation of the nucleic acids is followed by end repair of the nucleic acid fragments. End repair can include the generation of blunt ends, non-blunt ends (i.e. sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3'-end of the nucleic acid fragments, by a polymerase lacking 3'-exonuclease activity. End repair can be performed using any number of enzymes and/or methods known in the art

including, but not limited to, commercially available kits such as the Encore™ Ultra Low Input NGS Library System I. In some embodiments, end repair can be performed on double stranded DNA fragments to produce blunt ends wherein the double stranded DNA fragments contain 5' phosphates and 3' hydroxyls. In some embodiments, the double-stranded DNA fragments can be blunt-end polished (or "end-repaired") to produce DNA fragments having blunt ends, prior to being joined to adapters. Generation of the blunt ends on the double stranded fragments can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded DNA fragments can be blunt ended by the use of a single stranded specific DNA endonuclease, for example, but not limited to, mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, or any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded fragments generated by fragmenting the sample comprising nucleic acids. In still other cases, the nucleic acid fragments can be made blunt ended by filling in the overhanging single stranded ends of the double stranded fragments. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions

IDT Exhibit 1019.0099

of the double stranded fragments. Alternatively, the double stranded DNA fragments can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0044] U.S. Patent Publication Nos. 2013-0231253 A1 and 2014-0274729 A1 further describe methods of generating nucleic acid fragments, methods of modifying the fragments and analysis of the fragments, and are incorporated by reference in their entirety herein.

Ligation of Adaptors

[0045] Ligation of adaptors at the desired end of the sequence regions of interest (e.g., at the 5' or 3' end of a nucleic acid fragment generated form a sample) is suitable for carrying out the methods of the invention. Various ligation modalities are envisioned, dependent on the choice of nucleic acid, nucleic acid modifying enzymes and the resulting ligatable end of the nucleic acid. For example, when a blunt end product comprising the target region/sequence of interest is generated, blunt end ligation can be suitable. Alternatively, where the cleavage is carried out using a restriction enzyme of known sequence specificity, leading to the generation of cleavage sites with known sequence overhangs, suitable ends of the adaptors can be designed to enable hybridization of the adaptor to the cleavage site of the sequence region of interest and subsequent ligation. Ligation also refers to any joining of two nucleic acid molecules that results in a single nucleic acid sequence that can be further modified to obtain the sequence of the nucleic acids in question. Reagents and methods for efficient and rapid ligation of adaptors are commercially available, and are known in the art.

[0046] In some embodiments, the 5' and/or 3' end nucleotide sequences of fragmented nucleic acids are not modified or end-repaired prior to ligation with the adapter oligonucleotides of the present invention. For example, fragmentation by a restriction endonuclease can be used to leave

a predictable overhang, followed by ligation with one or more adapter oligonucleotides comprising an overhang complementary to the predictable overhang on a nucleic acid fragment. In another example, cleavage by an enzyme that leaves a predictable blunt end can be followed by ligation of blunt-ended nucleic acid fragments to adapter oligonucleotides comprising a blunt end. In some embodiments, end repair can be followed by an addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides, such as one or more adenine, one or more thymine, one or more guanine, or one or more cytosine, to produce an overhang. Nucleic acid fragments having an overhang can be joined to one or more adapter oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3' ends of end-repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adapters each having a thymine at a 3' end. In some embodiments, adapter oligonucleotides can be joined to blunt end double-stranded nucleic acid fragments which have been modified by extension of the 3' end with one or more nucleotides followed by 5' phosphorylation. In some cases, extension of the 3' end can be performed with a polymerase such as for example Klenow polymerase or any of the suitable polymerases provided herein, or by use of a terminal deoxynucleotide transferase, in the presence of one or more dNTPs in a suitable buffer containing magnesium. In some embodiments, nucleic acid fragments having blunt ends can be joined to one or more adapters comprising a blunt end. Phosphorylation of 5' ends of nucleic acid fragments can be performed for example with T4 polynucleotide kinase in a suitable buffer containing ATP and magnesium. The fragmented nucleic acid molecules may optionally be treated to dephosphorylate 5' ends or 3' ends, for example, by using enzymes known in the art, such as phosphatases.

[0047] In some embodiments, appending the adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction. In some embodiments, appendage of an adaptor to the nucleic acid fragments comprises ligation. In some embodiments, ligation of the adaptor to the nucleic acid fragments can be following end repair of the nucleic acid fragments. In another embodiment, the ligation of the adaptor to the nucleic acid fragments can be following generation of the nucleic acid fragments without end repair of the nucleic acid fragments. The adaptor can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors in which the adaptor comprises two complementary strands. In some embodiments, the adaptor can be a double stranded DNA adaptor. In some embodiments, the adaptor can be an oligonucleotide of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the adaptor is appended or attached. In some embodiments, the adaptor can be a conventional duplex adaptor, wherein the adaptor comprises sequence well known in the art. In a some embodiments, the adaptor can be appended to the nucleic acid fragments generated by the methods described herein in multiple orientations. In some embodiment, the methods described herein can involve the use of a duplex adaptor comprising double stranded DNA of known sequence that is blunt ended and can bind to the double stranded nucleic acid fragments generated by the methods described herein in one of two orientations. In some embodiments, the adaptor can be ligated to each of the nucleic acid fragments such that each of the nucleic acid fragments comprises the same adaptor. In other words, each of the nucleic acid fragments comprises a common adaptor. In another embodiment, an adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein such that each nucleic acid fragment in the library of nucleic acid

fragments comprises the adaptor ligated to one or both ends. In another embodiment, more than one adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein. The multiple adaptors may occur adjacent to one another, spaced intermittently, or at opposite ends of the nucleic acid fragments. In some embodiments, the adaptor can be ligated or appended to the 5' and/or 3' ends of the nucleic acid fragments generated by the methods described herein. The adaptor can comprise two strands wherein each strand comprises a free 3' hydroxyl group but neither strand comprises a free 5' phosphate. In some embodiments, the free 3' hydroxyl group on each strand of the adaptor can be ligated to a free 5' phosphate present on either end of the nucleic acid fragments of the present invention. In this embodiment, the adaptor comprises a ligation strand and a non-ligation strand whereby the ligation strand can be ligated to the 5' phosphate on either end of the nucleic acid fragment while a nick or gap can be present between the non-ligation strand of the adaptor and the 3' hydroxyl on either end of the nucleic acid fragment. In some embodiments, the nick or gap can be filled in by performing a gap repair reaction. In some embodiments, the gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In some embodiments, the gap repair can be performed using a DNA-dependent DNA polymerase with weak or no strand displacement activity. In some embodiments, the ligation strand of the adaptor can serve as the template for the gap repair or fill-in reaction. The gap repair or fill-in reaction may comprise an extension reaction wherein the ligation strand of the adaptor serves as a template and leads to the generation of nucleic acid fragments with complementary termini or ends. In some embodiments, the gap repair can be performed using Taq DNA polymerase. In some embodiments, the ligation of the first adaptor to the nucleic acid fragments generated by the

methods described herein may not be followed by gap repair. The nucleic acid fragments may comprise the adaptor sequence ligated only at the 5' end of each strand.

[0048] Ligation and, optionally gap repair, of the adaptor to the nucleic acid fragments generates an adaptor-nucleic acid fragment complex. In some embodiments, the adaptor-nucleic acid fragment complex can be denatured. Denaturation can be achieved using any of the methods known in the art including, but not limited to, physical, thermal, and/or chemical denaturation. In some embodiments, denaturation can be achieved using thermal or heat denaturation. In some embodiments, denaturation of the adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising the adaptor sequence at only the 5'end of the nucleic acid fragments. In another embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising adaptor sequence at both the 5'end and 3'end of the nucleic acid fragments.

Methods of Amplification

[0049] The methods, compositions and kits described herein can be useful to generate amplification-ready products directly from a nucleic acid source for downstream applications such as next generation sequencing, as well as generation of libraries with enriched population of sequence regions of interest. In some embodiments, the adapter-nucleic fragment ligated products, e.g., from the ligation of an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, is amplified.

[0050] Methods of amplification are well known in the art. In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal. In

some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR).

[0051] Suitable amplification reactions can be exponential or isothermal and can include any DNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), linear amplification, multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No, 6,251,639), Ribo-SPIA, or a combination thereof. In some cases, the amplification methods for providing the template nucleic acid may be performed under limiting conditions such that only a few rounds of amplification (e.g., 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 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[0052] PCR is an in vitro amplification procedure based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

[0053] LCR uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is

employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[0054] SDA (Westin et a 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3'-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[0055] Some aspects of the invention utilize linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

Downstream Applications

[0056] One aspect of the present invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as in next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. The methods of the present invention can also be used in the analysis of genetic information of selective genomic regions of interest (e.g., analysis of SNPs or other disease markers) as well as genomic regions which may interact with the selective region of interest. The methods of the present invention may further be used in the analysis of copy number variation as well as differential expression.

Sequencing

[0057] In some embodiments, a population of sequencing reads is generated from the amplified adapter-nucleic fragment ligated products. In some embodiments, a sequencing read comprises an index read, which comprises the sequence of an indexing site. In some embodiments, an index read comprises the sequence of an indexing site and the sequence of an identifier site. For example, the indexing site is sequenced with the identifier site. In some embodiments, the index read does not include the sequence of an identifier site. For example, the indexing site is not sequenced with the identifier site. In some embodiments, a sequencing read comprises the target sequence. In some embodiments, a sequencing read comprises a target sequence and an identifier sequence. For example, the target sequence is sequenced with the identifier site. In some embodiments, the target sequence is not sequenced with the identifier site.

[0058] The methods of the present invention are useful for sequencing by the method commercialized by Illumina, as described U.S. Pat Nos. 5,750,341; 6,306,597; and 5,969,119.

[0059] In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g., (A)/(A') or both ends (e.g., (A)/(A') and (C)/(C')). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g., by SPIA or linear PCR).The resulting nucleic acid is then denatured and the single-stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time.

[0060] In some embodiments, the methods of the invention are useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature* (2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711;

US20060024678; US20080213770; and US20080103058. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452; 7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764.

[0061] An example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single head, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

[0062] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g., Soni G V and Meller A. (2007) *Clin Chem*

53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

Data Analysis

[0063] In some embodiments, the sequence reads are used in the analysis of genetic information of selective genomic regions of interest as well as genomic regions which may interact with the selective region of interest. Amplification methods as disclosed herein can be used in the devices, kits, and methods known to the art for genetic analysis, such as, but not limited to those found in U.S. Pat. Nos. 6,449,562, 6,287,766, 7,361,468, 7,414,117, 6,225,109, and 6,110,709.

[0064] In some embodiments, the sequencing reads are used to detect duplicate sequencing reads. In some embodiments, a sequencing read is identified as a duplicate sequencing read when it contains an identifier site and target sequence that is the same as another sequencing read from the same population of sequencing reads.

[0065] In some embodiments, duplicate sequencing reads are differentiated from one another as being true duplicates versus apparent or perceived duplicates. Apparent or perceived duplicates may be identified from sequencing libraries and using conventional measures of duplicate reads (i.e., reads were mapped using bowtie), where all reads with the same start and end nucleic acid coordinates were counted as duplicates. True duplicates may be identified from sequencing

libraries having had an identifier site introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates.

[0066] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acids generated from any dsDNA may have the same start site, as determined from the sequencing reads of the target sequence of the generated nucleic acid fragments. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original dsDNA molecule and, therefore, are no true duplicate reads. The ligation of random sequences onto dsDNA molecules, accompanied by the methods of the invention, allow for the identification of true duplicate reads versus apparent or perceived duplicate reads.

[0067] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acid fragments generated from a genomic DNA (gDNA) molecule that have the same start site, which can be determined from the sequencing reads of the target sequence of the nucleic acid fragments is determined. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original gDNA molecule and, therefore, are not true duplicate reads. In another embodiment, an identifier site is inserted at the adapter insert junction. The sequence of the identifier site is carried through the library amplification step. The identifier site is the first sequence read during the forward read. As the identifier sequence is not logically present adjacent to naturally occurring sequence, this uniquely identifies the DNA fragment. Therefore, by ligating random sequences onto the original gDNA, the methods of the invention identify true duplicate reads.

[0068] In some embodiments, a duplicate sequencing read is detected and analyzed. Duplicate reads can be filtered using 'samtools rmdup', wherein reads with identical external coordinates

are removed, only one read with highest mapping quality is retained. After filtering, a filtered set of deduplicated reads can be used in any downstream analysis. Conversely, this filtering step can be skipped, and downstream analysis can be done using the unfiltered reads, including duplicates.

[0069] In some embodiments, sequencing reads are generated from a number of samples. In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from the samples, in which the nucleic acid fragments from each sample has the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before sequencing reads of the target sequence and/or identifier site are analyzed. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before duplicate sequencing reads are analyzed and/or removed.

[0070] In some embodiments, a method disclosed herein identifies or detects one or more true duplicate(s) with increased accuracy as compared to other methods. For example, in some embodiments, a method disclosed herein identify true duplicates (in contrast to identifying apparent or perceived duplicates) with increased accuracy as compared to other methods. The increased resolution and/or accuracy in identifying one or more true duplicate(s) can provide a considerable contribution to the state of the art in more accurately identifying true duplicates. In some embodiments, a method disclosed herein identifies or detects a true duplicate with

increased efficiency as compared to other methods, such as paired end sequencing. The increase in accuracy, resolution, and/or efficiency in detecting duplicate reads (e.g., true duplicate(s)) can increase confidence in sequencing results, such as for expression and CNV analysis.

Kits

[0071] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for amplification.

[0072] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[0073] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit may be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[0074] A kit may include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

[0075] In some embodiments, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, a kit comprises a composition of the invention, in one or more containers. In some embodiments, the

invention provides kits comprising adapters, primers, and/or other oligonucleotides described herein. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) a forward adapter (e) one or more oligonucleotides comprising reverse adaptor sequence and (f) one or more buffers suitable for one or more of the elements contained in said kit. The adapters, primers, other oligonucleotides, and reagents can be, without limitation, any of those described above. Elements of the kit can further be provided, without limitation, in any of the amounts and/or combinations (such as in the same kit or same container) described above. The kits may further comprise additional agents, such as those described above, for use according to the methods of the invention. For example, the kit can comprise a first forward adaptor that is a partial duplex adaptor as described herein, a second forward adapter, and a nucleic acid modifying enzyme specific for a restriction and/or cleavage site present in the first forward adaptor. The kit elements can be provided in any suitable container, including but not limited to test tubes, vials, flasks, bottles, ampules, syringes, or the like. The agents can be provided in a form that may be directly used in the methods of the invention, or in a form that requires preparation prior to use, such as in the reconstitution of lyophilized agents. Agents may be provided in aliquots for single-use or as stocks from which multiple uses, such as in a number of reaction, may be obtained.

[0076] In some embodiments, the kit comprises a plurality of adaptor oligonucleotides, wherein each of the adaptor oligonucleotides comprises at least one of a plurality of identifier site sequences, wherein each identifier site sequence of the plurality of identifier site sequences differs from every other identifier site sequence in said plurality of identifier site sequences at at least three nucleotide positions, and instructions for using the same. Adapters comprising

different identifier site sequences can be supplied individually or in combination with one or more additional adapters having a different identifier site sequence. In some embodiments, the kit can comprises a plurality of adapter oligonucleotides.

EXAMPLES

Example 1: Identification of Duplicate Sequencing Reads with NuGEN Ovation Target Enrichment Library System

[0077] Sample Description: 100ng of DNA from a human HapMap sample (NA19238) was fragmented to approximately 500 base pair in length by sonication with a Covaris system (Covaris, Inc., Woburn, MA). The resulting DNA was treated with end repair enzyme mix NuGEN R01280 and R01439 (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendation to produce blunt ended DNA fragments.

[0078] Library Generation. Enrichment, and Identifier Site Incorporation: An oligonucleotide with segments, from 5' to 3' of the top strand, 1) an Illumina indexing read priming site such as AGAGCACACGTCTGAACTCCAGTCAC (SEQ ID NO:2), 2) an indexing site, 3) an identifier site having a random 6 base sequence and, 4) a sequence compatible with an Illumina forward sequencing priming site such as TCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO:3) , was annealed to a second oligonucleotide to form a partially double stranded DNA adapter. Five uM of these adapters were ligated onto the end-repaired DNA using Ligase and Ligase reaction buffer from NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendations. Following 30 minutes of incubation at 25°C, the reaction mixture was diluted with water, 0.8X volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated

DNA fragments eluted according to manufacturer's recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95°C then slowly cooling the mixture from 80°C to 60°C by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, MA) according to the manufacturer's protocols. Following extension, the DNA fragments were collected on Agencourt magnetic beads, washed and eluted according to manufacturer's recommendations. These libraries were enriched by 30 cycles of PCR using NuGEN library enrichment primers (Ovation Target Enrichment Library System, NuGEN Technologies, Inc., San Carlos, CA) that also contain the Illumina flow cell sequences (Illumina Inc., San Diego, CA) according to supplier's recommendation.

[0079] The resulting libraries were quantitated by qPCR using a kit provided by KAPA, diluted to 2nM and applied to an Illumina MiSeq DNA Sequencer (Illumina Inc., San Diego, CA). The following series was run: 36 base first read, 14 base second read, and 24 base third read.

[0080] *Data Analysis:* The sequencer output was processed in accordance with manufacturer's recommendation. In order to analyze the data, the indexing read was split into two files. The first file contained the first 8 bases of the indexing reads and is utilized as the library index file for standard library parsing. The other file contains only the random bases and is set aside for further sequence parsing.

[0081] Following our data analysis pipeline of sequence alignment with bowtie aligner (Langmead B. et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009, 10:R2.), duplicate reads were identified by their genomic start positions. At this point, sequencing reads that started at the same genomic position were checked against the random bases file to see if they have the same or different set of random

bases that were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation Target Enrichment targeting probe was used to generate the sequence fragment in question. These sequences, therefore, did not provide unique information about the starting genomic DNA and are considered as one sequencing read for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. **Figure 5** provides the analysis results demonstrating the identification of the duplicate reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

[0082] If the sequences of two libraries are identical, their duplication status is unknown since this could occur by chance in any library. If an identifier sequence is used in combination with the library reads, the status can be determined (duplicates if identical, distinct if different). With the SPET system, one end is common so the probability of two libraries having identical ends increases. These would appear to be duplicate sequences and their true status can be determined by looking at an identifier sequence.

[0083] Over the sampling of all randomly selected reads, the use of identifier sites provided an increased resolution on the presence of true duplicates. When evaluating two million random reads, the apparent duplicates were found to comprise 39% of all reads. However, the true duplicates, identified through the use of an identifier site, were found to comprise only 26% of all reads. The methods employing the use of an identifier site were found to considerably increase the resolution of the true number of duplicates within the pool of reads.

Example 2: Removal of Duplicate Sequencing Reads with 8 Base Identifier Site

[0084] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0085] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites. This identifier site is designed as eight random bases introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain the 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rndup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 3: Removal of Duplicate Sequencing Reads with Mixture of Random 1-8 Base Identifier Sites

[0086] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the

ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0087] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0088] One to eight random bases are introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain between 1 and 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rmdup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 4: Removal of Duplicate Sequencing Reads with Mixture of 96 Defined 6-Base Identifier Sites

[0089] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0090] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0091] A mixture of 96 defined six-bases sequences are introduced onto the end of the adaptors. Thus, each six-base sequence is an identifier site. Sequence reads from libraries made with such adaptors contain one of the 96 six-base identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rmdup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 5: Identification of Duplicate Sequencing Reads in Determining mRNA Expression Levels

[0092] Sample Description: Total RNA is extracted from tumor and normal adjacent tissue for the purpose of finding differences in expression levels of transcripts between the two sample types. l00ng of each sample is converted into cDNA using the USP primers, reaction buffer, and Reverse Transcriptase provided in NuGEN's Encore Complete Library System (NuGEN Technologies, Inc., San Carlos, CA) according to the supplier's recommendations. This is followed by second strand synthesis, again using materials provided in the kit according to recommendations. Double stranded cDNA was prepared using the Superscript® Double-Stranded cDNA Synthesis Kit from Life Technologies (Carlsbad, CA) according the manufacturer's instructions. DNA was sheared with a Covaris S-series device (Covaris, Inc., Woburn, MA) using the 200 bp sonication protocol provided with the instrument (10% duty cycle, 200 cycles/burst, 5 intensity, 180 seconds). DNA was treated with 1.5 uL 10X Blunting Buffer, 0.5 uL Blunting Enzyme (New England Biolabs, Inc., Ipswich, MA; p/n E1201) and 1.2 uL of 2.5mM of each dNTP mix in a total volume of 15 uL for 30 minutes at 25°C followed by 10 minutes at 70°C.

[0093] *Library Generation*: The DNA fragments were then subjected to end repair using end repair buffers and enzymes provided in NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA).

[0094] Forward adaptor 5'

[0095] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTNNNNNNNN (SEQ ID NO:4),

[0096] Reverse adaptor 1) 5'

[0097] CAAGCAGAAGACGGCATACGAGATTCCCTTGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCTNNNNNNNN (SEQ ID NO:5),

[0098] Reverse adaptor 2) 5'

[0099] CAAGCAGAAGACGGCATACGAGATTGAAGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCNNNNNNNN (SEQ ID NO:6), and

[0100] Common partner 5' NNNNNNNNAGATCGGAAGAGC (SEQ ID NO:7) were all ordered from IDT (Integrated DNA Technologies, Coralville, IA). The reverse adaptors each contain a unique identifier (underlined) that enables libraries made with these adapters to be distinguished. In this case N represents an equimolar mixture of A, C, G, and T. A mixture of 5uM forward, 5uM reverse, and l0uM common in l0mM MgCl2, 50mM Tris pH 8 was heated to 95C for 5 minutes, then cooled to 20C. Adaptor ligation was performed by addition of 4.5 uL water, 3 uL Adaptor mix (prepared above), 6 uL 5X NEBNext Quick Ligation Reaction Buffer and 1.5 uL Quick T4 DNA Ligase (New England Biolabs, Inc., Ipswich, MA; p/n E6056), followed by incubation for 30 minutes at 25°C followed by 10 minutes at 70°C. Ligation products were purified by adding 70uL water and 80 uL of Ampure XP beads (Agencourt Genomics), washing twice with 70% ethanol and eluting with 20 uL of l0mM Tris pH 8.0.

Library products were amplified in a 50uL PCR containing 0.5uM each of primer (5' AATGATACGGCGACCACCGA (SEQ ID NO:8), and 5' CAAGCAGAAGACGGCATACGA (SEQ ID NO:9), 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM MgCl2, 0.2mM each dNTP, and 1 unit Taq polymerase. The reaction was cycled 15 times under the conditions 95C for 15 seconds, 60C for 1 minute. PCR products were purified with 1 volume of Ampure XP beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) as described above. The library was analyzed by HS DNA Bioanalyzer (Agilent Technologies, Santa Clara, CA) and quantitated with the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA; p/n KK4835) according to the supplied instructions. The resulting libraries are combined and are compatible with standard TruSeq single end or paired Illumina sequencing protocols for GAIIx, MiSeq, or Hi Seq sequencing instruments (Illumina Inc., San Diego, CA). The following series is run: 50 base first read, 6 base second read. A third read is not required for counting purposes or duplication analysis.

[0101] Data Analysis: The sequencer output was processed in accordance with manufacturer's recommendation. The 6 bases of the index read is used for standard library parsing, separating the data files from the two sample types. The first 50 bases of the target sequence read are compared to each other. Any read that has identical sequence to any other read is identified as a duplicate and removed from the population, thus, only a single copy is retained within the file. Once the duplicate reads have been removed, 8 bases are trimmed from each read. The trimmed reads are aligned to a reference genome. Differential expression is then determined by comparing FPKM (fragments per kilobase per million reads) values between libraries utilizing scripts such as cufflinks or cuffdiff (Trapnell et al. 2010, Nature Biotechnology, 28, 511-515; Trapnell et al. 2013, Nature Biotechnology, 31, 46-53).

Example 6: Sequencing of the Identifier Site With the Target Sequence in Paired End Reads

[0102] The Ovation Library System for Low Complexity Samples (NuGEN Technologies, Inc., San Carlos, CA) was used to generate four libraries, each from a single amplicon, following manufacturer's protocol. Purified libraries were mixed and sequenced as a multiplex on the Illumina MiSeq (Illumina Inc., San Diego, CA) to produce 125nt forward, 8nt index 1, 8nt index 2, and 25nt reverse reads. Because all amplicon reads start and end at the same sequence coordinates, the traditional method of marking library PCR duplicates (marking reads that start and end at the same genomic coordinates as duplicates) cannot be used. Instead, the 0-8nt of random sequence contained in the adaptors ligated to the amplicon were treated as an identifier sequence and used to mark duplicates. Any paired end reads that shared the same length and sequence of these random bases with any other paired end read was called a duplicate. The table below shows the results of this duplicate marking.

Table 1.

Library	Reads	Duplicate Reads	Reads from independent molecules
1	454249	376797	77452
2	439367	364001	75366
3	317760	253057	64703
4	476572	398380	78192

[0103] Table 1 demonstrates the accuracy of the method used to differentiate between duplicate reads and reads from truly independent molecules. The population of reads from independent molecules, depicted in the final column of Table 1, represent sequences from independent amplicon molecules used in generating the libraries.

Example 7: Sequencing of the Identifier Site With the Target Sequence in Reduced Representation Bisulfite (RRBS) Libraries

[0104] Reduced representation bisulfite (RRBS) libraries of the human genome are generated through the complete restriction enzyme digestion of 100ng input sample, followed by selection for short fragments. The resulting pool of fragments are ligated to adaptor sequences comprising an indexing site and an identifier site. The identifier sites comprise either 6 or 8 random nucleotides. The sequences are then sequenced to identify the identifier sites, thus revealing the true number of duplicates in the pool. In the absence of identifier sites, the number of apparent or perceived duplicates are greater than the number of true duplicates. The inclusion of an identifier site results in the identification of the number of true duplicates, as compared to the larger number of apparent or perceived duplicates.

[0105] Unless defined otherwise, all technical and scientific terms herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials, similar or equivalent to those described herein, can be used in the practice or testing of the present invention, the preferred methods and materials are described herein. All publications, patents, and patent publications cited are incorporated by reference herein in their entirety for all purposes.

[0106] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

[0107] All references, articles, publications, patents, patent publications, and patent applications cited herein are incorporated by reference in their entireties for all purposes. However, mention of any reference, article, publication, patent, patent publication, and

patent application cited herein is not, and should not be taken as an acknowledgment or any form of suggestion that they constitute valid prior art or form part of the common general knowledge in any country in the world.

[0108] While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth and as follows in the scope of the appended claims.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	NUGEN TECHNOLOGIES, INC.	ART UNIT:	Unassigned
SERIAL NUMBER:	16/743,604	CONF. NO.:	3249
FILING DATE:	January 15, 2020	EXAMINER:	Unassigned
TITLE:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		

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

FILED ELECTRONICALLY

AMENDMENT IN RESPONSE TO NOTICE

This paper is in response to the Notice to Comply with Requirements for Patent Applications Containing Nucleotide and/or Amino Acid Sequence Disclosures mailed February 18, 2020 from the United States Patent and Trademark Office.

Amendments to the Specification begin on page 2.

A **Sequence Listing** is referred to on page 3.

Amendments to the Claims begin on page 4.

Remarks begin on page 6.

Applicant believes that no fees are due in connection with this Amendment. However, Applicant authorizes any required fees to be charged to Deposit Account No. 500369.

Electronic Acknowledgement Receipt

EFS ID:	39188096
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	17-APR-2020
Filing Date:	15-JAN-2020
Time Stamp:	13:00:51
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	Sequence Listing (Text File)	NUGN-001-03US_SeqList_ST25.TXT	2694	no	-

Warnings:

Information:					
2	Applicant Arguments/Remarks Made in an Amendment	NUGN-001-03US_Spec_Marked.PDF	321115 a7fc28464f86ded182d2e3ffb08790685c9911be	no	40
Warnings:					
Information:					
3		NUGN-001-03US_SeqList_Amendment.PDF	92004 69f0aca5c5c080658117944231a6fd4ea3a9c855	yes	7

Multipart Description/PDF files in .zip description

Document Description	Start	End
Response to Pre-Exam Sequence Notice	1	1
Specification	2	2
Sequence Listing	3	3
Claims	4	5
Applicant Arguments/Remarks Made in an Amendment	6	7

Warnings:					
Information:					
4	Specification	NUGN-001-03US_Spec_Clean.PDF	320524 e5a67332c879f8da8a9700eb9717220f6e503e01	no	40
Warnings:					
Information:					
		Total Files Size (in bytes):	736337		

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.

Amendments to the Claims

Pursuant to 37 C.F.R. §1.121(c), this listing of claims will replace all prior versions and listings of claims in this application.

1. - 21. (Cancelled)

22. (New) A method for detecting duplicate sequencing reads, the method comprising:

ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides and a primer binding site;

amplifying the adaptor-ligated fragments into amplicons;

sequencing the amplicons to produce sequence reads that include identifier and target sequences; and

identifying sequence reads with identical identifier and target sequences as duplicates.

23. (New) The method of claim 22, further comprising removing a duplicate read from the sequence reads.

24. (New) The method of claim 22, wherein the identifier site is sequenced with the target sequence.

25. (New) The method of claim 22, wherein the identifier site is sequenced separately from the target sequence.

26. (New) The method of claim 22, wherein the nucleic acid fragments are comprise RNA.

27. (New) The method of claim 26, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

IDT Exhibit 1019.0130

28. (New) The method of claim 22, wherein the primer binding site is a universal target sequencing primer binding site.

29. (New) The method of claim 22, wherein each unique identifier is 6 nucleotides in length.

30. (New) The method of claim 22, wherein each unique identifier is 8 nucleotides in length.

31. (New) The method of claim 22, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.

32. (New) The method of claim 22, further comprising determining the genomic start position for each sequencing read.

IDT Exhibit 1019.0131

Remarks

By this amendment, claims 1-21 have been cancelled and claims 22-32 added, with claim 22 being the sole independent claim. Support for new claims 22-32 can be found throughout the subject application as originally filed, including at paragraphs [0008], [0013], [0026], [0030]-[0031], [0045], and [0064]-[0070].

No new matter is added.

Sequence Listing

The February 18, 2020 Notice to Comply with Requirements for Patent Applications Containing Nucleotide and/or Amino Acid Sequence Disclosures (Notice) requires, pursuant to 37 C.F.R. §1.821(g), the submission of a sequence listing in compliance with 37 C.F.R. §§1.821-1.825.

Applicant is filing herewith a Sequence Listing in computer readable form (CRF) in the ASCII text file named: NUGN-001-03US_SeqList_ST25.txt. The Sequence Listing is 3 kilobytes in size, was created on April 3, 2020 in the IBM-PC machine format, and has an operating system compatibility with MS-Windows.

In accordance with 37 C.F.R. §1.821(g), the Sequence Listing includes no new matter. The sequences were found and identified in the as-filed application, including at paragraphs [0020], [0078], and [0095]-[0100].

In accordance with MPEP §2422.03(a), the ASCII text file of the Sequence Listing will serve as both the paper copy and the computer readable form (CRF) of the Sequence Listing. Accordingly, Applicant is not submitting herewith: (1) a second copy of the sequence listing as a PDF file; (2) a statement under 37 C.F.R. §1.821(f) indicating that the paper copy and CRF of the sequence listing are identical; or (3) a request to use a compliant computer readable form of the sequence listing that is already on file for another application pursuant to 37 C.F.R. §1.821(e).

IDT Exhibit 1019.0132

The Notice required submission of a substitute specification in compliance with 37 C.F.R. §1.121(b)(3) and §1.125. In accordance with 37 C.F.R. §1.121(b)(3) and §1.125. Applicant is filing herewith a substitute specification that has been amended to refer to the Sequence Listing.

In accordance with 37 C.F.R. §1.125(c), Applicant is also submitting herewith a copy of the substitute specification with markings showing changes necessary to derive the current specification, including those changes made in the January 15, 2020 Preliminary Amendment filed as part of the subject application.

The substitute specification includes no new matter.

Applicant submits that the subject application complies with the requirements of 37 C.F.R. §§1.821-1.825.

If there are any questions concerning this application, please contact Applicant's undersigned representative.

Dated: April 17, 2020

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: 617-856-8483

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

63724082

IDT Exhibit 1019.0133

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPFICATE SEQUENCING READ

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. Patent Application Serial No. 15/406,002, filed January 13, 2017, which application is a Division of U.S. Application 14/540,917, filed November 13, 2014, now issued as U.S. Pat. 9,546,399, each of which claims the benefit of U.S. Provisional Application Serial No. 61/903,826, filed November 13, 2013, each of which is herein incorporated by reference in its entirety.

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0002] The contents of the text file submitted electronically herewith are incorporated herein by reference in their entirety: A computer readable format copy of the Sequence Listing (filename: NUGN-001-03US_SeqList_ST25.txt, date recorded: April 3, 2020, file size 3 kilobytes).

FIELD OF THE PRESENT INVENTION

[0003] The present invention relates generally to the field of high throughput sequencing reactions and the ability to discern artifacts arising through sequence duplications from nucleotide molecules that are unique molecules.

BACKGROUND OF THE PRESENT INVENTION

[0004] In RNA sequencing applications, accurate gene expression measurements may be hampered by PCR duplicate artifacts that occur during library amplification. When analyzing RNA sequencing data, when two or more identical sequences are found, it can be difficult to know if these represent unique cDNA molecules derived independently from different RNA molecules, or if they are PCR duplicates derived from a single RNA molecule. In genotyping by

sequencing, duplicate reads can be considered non-informative and may be collapsed down to a single read, thus reducing the number of sequencing reads used in final analysis. Generally, sequencing reads may be determined to be duplicates if both forward and reverse reads have identical starting positions, even though two independently generated molecules can have identical starting positions by random chance. Single primer extension based targeted re-sequencing suffers from an issue in that only one end of a sequencing read is randomly generated, while the other (reverse read) end is generated by a specific probe. This may make it difficult to determine if two reads are duplicates because they have been duplicated by PCR or because by chance they happened to start at the same position.

[0005] In expression analysis studies there may be limited value in doing paired end sequencing since the goal of the experiment is to determine amounts of transcript present as opposed to studying exon usage. In these studies, paired end sequencing adds costs while the only value is in helping distinguish PCR duplicates. The probability of two reads starting in the same position on only one end is higher than the probability of two reads having the same starting position on two ends (forward and reverse read). There is a need for improved methods that allow for low-cost, high throughput sequencing of regions of interest, genotyping or simple detection of RNA transcripts without inherent instrument inefficiencies that drive up sequencing costs due to the generation of unusable or non-desired data reads. The invention described herein fulfills this need. Here, we describe an adaptor approach that allows for the identification of true PCR duplicates and their removal.

[0006] The methods of the present invention provide novel methods for identifying true duplicate reads during sequencing, such as to improve data analysis of sequencing data, and other related advantages.

SUMMARY OF THE PRESENT INVENTION

[0007] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The detection and/or removal of duplicate sequencing reads presented herein is a novel approach to increasing the efficacy of evaluating data generated from high throughput sequence reactions, including complex multiplex sequence reactions.

[0008] Accordingly, the present invention provides a method of detecting a duplicate sequencing read from a population of sample sequencing reads, the method comprising ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site. The ligated adaptor-nucleic acid fragment products can be amplified, thus generating a population of sequencing reads from the amplified adaptor-nucleic acid ligation products. The sequencing reads with a duplicate identifier site and target sequence can then be detected from the population of sequencing reads. The methods can further include the removal of the sequencing reads with the duplicate identifier site and target sequence from the population of sequence reads.

[0009] In some embodiments, the identifier site is sequenced with the indexing site or the target sequence. In further embodiments, the identifier site is sequenced separately from the indexing site or the target sequence.

[0010] In some embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the identifier site; and the target sequence primer binding site. In further embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the target sequence primer binding site; and the identifier site.

[0011] In some embodiments, the plurality of nucleic acid fragments is generated from more than one sample. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, the sequencing reads are separated based on the indexing site. In yet other embodiments, the separation of sequencing reads is performed prior to detecting sequence reads with a duplicate identifier site and target sequence.

[0012] In some embodiments, the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments. In further embodiments, the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

[0013] In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In further embodiments, the indexing site is about 6 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In further embodiments, the identifier site is about 8 nucleotides in length.

[0014] In some embodiments, the indexing primer binding site is a universal indexing primer binding site; and in some embodiments, the target sequence primer binding site is a universal target sequence primer binding site.

[0015] The present invention also encompasses embodiments that include a kit comprising a plurality of adaptors, wherein each adaptor comprise an indexing primer binding site; an indexing site, and identifier site, and a target sequencing primer binding site.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] A better understanding of the novel features of the invention and advantages of the present invention will be obtained by reference to the following description that sets forth

illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0017] Figure 1 depicts a schematic of generating sequencing reads of a library including where an indexing primer and a target sequence primer anneal.

[0018] Figure 2A depicts a mechanism of the Single Primer Enrichment Technology (SPET) and how an identifier site is carried over into the final library and how sequencing through the indexing site into the identifier site provides data on which nucleic acid molecule is being identified.

[0019] Figure 2B is a continuation of Figure 2A.

[0020] Figure 3 offers a detailed view of the sequence of an example adaptor, among many envisioned embodiments, and the position of the indexing and identifier sites (SEQ ID NO: 1). "N" refers to any nucleic acid.

[0021] Figure 4 depicts a schematic of two separate sequence libraries, indicating where the indexing primers and the target primers anneal in a traditional library as compared to a library using an identifier site.

[0022] Figure 5 depicts a data table demonstrating the accuracy of resolving true duplicates versus apparent or perceived duplicates using an identifier site in the adaptors.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0023] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The present invention encompasses methods of detecting sequences duplicated in sequencing applications, and further removal of the duplicated sequence reads. The present invention further encompasses kits

comprising components that would allow for customized applications of the method of detecting and removing duplicated sequence reads in high throughput sequencing reactions. The compositions and methods can be used with various applications for genetic sample analysis, such as RNA sequence analysis, copy number variation analysis, methylation sequencing analysis, genotyping and whole genome amplification.

[0024] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the present invention. On the contrary, the present disclosure is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the present invention.

[0025] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg 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.

[0026] In some embodiments, the methods disclosed herein is for detecting from a population of sequencing reads a sequencing read, such as a duplicate sequencing read with a duplicate identifier site and target sequence. A duplicate sequencing read can be a sequencing read with

the same identifier site and target sequence as another sequencing read in the population of sequencing reads.

Adaptor

[0027] The present invention provides compositions of adaptors and methods comprising use of an adaptor. An adapter refers to an oligonucleotide sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. The target polynucleotide molecules may be fragmented or not prior to the addition of adaptors. In some embodiments, a method disclosed herein comprises ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples.

[0028] Various adaptor designs are envisioned which are suitable for generation of amplification-ready products of target sequence regions/strands of interest. For example, the two strands of the adaptor may be self-complementary, non-complementary or partially complementary. In some embodiments, the adaptor can comprise an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site.

[0029] An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site. An indexing site is a nucleic acid sequence that acts as an index for multiple polynucleotide samples, thus allowing for the samples to be pooled together into a single sequencing run, which is known as multiplexing. In some embodiments, the indexing site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In some embodiments, the indexing site is about 6 nucleotides in length.

[0030] An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads. In some embodiments, the identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In some embodiments, the identifier site is about 8 nucleotides in length. This identifier site can be designed as a set of sequences, or it can be semi-random, or it can be completely random. In addition, this identifier site can be a fixed length, or it can be a variable length. In some embodiments, the identifier sites in a plurality of adaptors are of a fixed length. For example, the identifier sites can all be eight random bases. In another embodiment, the identifier sites in a plurality of adaptors are of a variable length. For example, the identifier sites can range in size from 1 to 8 bases. In yet another embodiment, the identifier sites can be of a defined set of defined sequence. For example, the identifier site of a plurality of adaptors can be one of 96 defined six-base nucleotide sequences.

[0031] A target sequence primer binding site nucleotide sequence for binding a primer for a target sequence. The primer can be used to amplify the target sequence (e.g., a nucleic acid fragment from a sample). Accordingly, in some embodiments, the adaptor comprises an indexing primer binding site and a target sequence primer binding site.

[0032] A primer is a polynucleotide chain, typically less than 200 residues long, most typically between 15 and 100 nucleotides long, but can encompass longer polynucleotide chains. A primer targeting the primer binding sites are typically designed to hybridize to single-stranded nucleic acid strands. In some embodiments, the primers targeting the primer binding sites are designed to hybridize to single-stranded DNA targets. In the case where the sample comprises genomic DNA or other double-stranded DNA, the sample can be first denatured to render the target single stranded and enable hybridization of the primers to the desired sequence regions of

interest. In these embodiments, the methods and compositions described herein can allow for region-specific enrichment and amplification of sequence regions of interest. In some embodiments, the other double-stranded DNA can be double-stranded cDNA generated by first and second strand synthesis of one or more target RNAs.

[0033] In other embodiments, the primers targeting the primer binding sites are designed to hybridize to double-stranded nucleic acid targets, without denaturation of the double stranded nucleic acids. In other embodiments, the primers targeting the primer binding sites are designed to hybridize to a double-stranded DNA target, without denaturation of the dsDNA. In these embodiments, the primers targeting the selected sequence regions of interest are designed to form a triple helix (triplex) at the selected sequence regions of interest. The hybridization of the primers to the double-stranded DNA sequence regions of interest can be carried out without prior denaturation of the double stranded nucleic acid sample. In such embodiments, the methods and compositions described herein can allow for region-specific enrichment as well as strand-specific enrichment and amplification of sequence regions of interest. This method can be useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and analysis of multiplicity of sequence regions of interest in the native complex nucleic acid sample. The method can find use for studies and analyses carried out in situ, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as permit the analysis of complex genomic DNA without disruption of chromatin structures.

[0034] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as for amplification of a target sequence (e.g., in PCR). The oligonucleotide primer can be a synthetic polynucleotide

that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a sequence of the target polynucleotide. In some embodiments, the 3' region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a primer binding site.

[0035] In some embodiments, the primer binding site is a binding site for a universal primer. A universal primer is a primer that can be used for amplifying a number of different sequences. In some embodiments, a universal primer is used to amplify different libraries. In some embodiments, an indexing primer binding site is a binding site for a universal indexing primer (i.e., the indexing primer binding site is a universal indexing primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal indexing primer binding site. In some embodiments, the universal indexing primer can be used to amplify and/or sequence a number of different indexing sites.

[0036] In some embodiments, a target sequence primer binding site is a binding site for a universal target sequence primer (i.e., the target sequence primer binding site is a universal target sequence primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal target sequence primer binding site. In some embodiments, the universal target sequence primer can be used to amplify and/or sequence a number of different target sequences.

[0037] In some embodiments, the adaptor comprises an identifier site 3' to an indexing site. In some embodiments, the adaptor comprises an identifier site 5' to an indexing site. In some embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target sequence primer binding site. In other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target

sequence primer binding site. In yet other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, identifier site, indexing site, and target sequence primer binding site.

Samples

[0038] In some embodiments, an adaptor is ligated to a nucleic acid fragment (e.g., the 5' end of the nucleic acid fragment). The nucleic acid fragment can be from a plurality of nucleic acid fragments from one or more samples. The nucleic acid fragment can be RNA, DNA, or complex DNA, for example genomic DNA and PNA, in which case one might use a modified nucleic acid. The nucleic acid fragment may also be cDNA. The cDNA can be generated from RNA, e.g., mRNA.

[0039] The sample can be a biological sample. For example, the sample can be an animal, plant, bacterial, algal, or viral sample. In some embodiments, the sample is a human, rat, or mouse sample. The sample can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The sample can be cDNA made from a mixture of genomes of different species. In some embodiments, the sample can be from a synthetic source. The sample can be mitochondrial DNA. The sample can be cell-free DNA. The cell-free DNA can be obtained from sources such as a serum or a plasma sample. The sample can comprise one or more chromosomes. For example, if the sample is from a human, the sample can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. In some embodiments, the sample comprises a linear or circular genome. The sample can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The sample can be from more than one individual or organism. The sample

can be double-stranded or single-stranded. The sample can be part of chromatin. The sample can be associated with histones.

[0040] In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from more than one sample, such as 2, 3, 4, 5, or more samples. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site.

[0041] In some embodiments, a population of nucleic acid fragments generated from a sample is of one or more specific size range(s). In some embodiments, the fragments have an average length from about 10 to about 10,000 nucleotides. In some embodiments, the fragments have an average length from about 50 to about 2,000 nucleotides. In some embodiments, the fragments have an average length from about 100-2,500, 10-1,000, 10-800, 10-500, 50-500, 50-250, or 50-150 nucleotides. In some embodiments, the fragments have an average length less than 10,000 nucleotide, such as less than 5,000 nucleotides, less than 2,500 nucleotides, less than 2,500 nucleotides, less than 1,000 nucleotides, less than 500 nucleotides, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides.

[0042] In some embodiments, fragmentation of the nucleic acids can be achieved through methods known in the art. Fragmentation can be achieved through physical fragmentation methods and/or enzymatic fragmentation methods. Physical fragmentation methods can include

nebulization, sonication, and/or hydrodynamic shearing. In some embodiments, the fragmentation can be accomplished mechanically comprising subjecting the nucleic acids in the input sample to acoustic sonication. In some embodiments, the fragmentation comprises treating the nucleic acids in the input sample with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of nucleic acid or polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g., from New England Biolabs). For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some embodiments, fragmentation comprises treating the nucleic acids in the input sample with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel.

[0043] In some embodiments, fragmentation of the nucleic acids is followed by end repair of the nucleic acid fragments. End repair can include the generation of blunt ends, non-blunt ends (i.e. sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3'-end of the nucleic acid fragments, by a polymerase lacking 3'-exonuclease activity. End repair can be performed using any number of enzymes and/or methods known in the art

including, but not limited to, commercially available kits such as the Encore™ Ultra Low Input NGS Library System I. In some embodiments, end repair can be performed on double stranded DNA fragments to produce blunt ends wherein the double stranded DNA fragments contain 5' phosphates and 3' hydroxyls. In some embodiments, the double-stranded DNA fragments can be blunt-end polished (or "end-repaired") to produce DNA fragments having blunt ends, prior to being joined to adapters. Generation of the blunt ends on the double stranded fragments can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded DNA fragments can be blunt ended by the use of a single stranded specific DNA endonuclease, for example, but not limited to, mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, or any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded fragments generated by fragmenting the sample comprising nucleic acids. In still other cases, the nucleic acid fragments can be made blunt ended by filling in the overhanging single stranded ends of the double stranded fragments. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions

of the double stranded fragments. Alternatively, the double stranded DNA fragments can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0044] U.S. Patent Publication Nos. 2013-0231253 A1 and 2014-0274729 A1 further describe methods of generating nucleic acid fragments, methods of modifying the fragments and analysis of the fragments, and are incorporated by reference in their entirety herein.

Ligation of Adaptors

[0045] Ligation of adaptors at the desired end of the sequence regions of interest (e.g., at the 5' or 3' end of a nucleic acid fragment generated form a sample) is suitable for carrying out the methods of the invention. Various ligation modalities are envisioned, dependent on the choice of nucleic acid, nucleic acid modifying enzymes and the resulting ligatable end of the nucleic acid. For example, when a blunt end product comprising the target region/sequence of interest is generated, blunt end ligation can be suitable. Alternatively, where the cleavage is carried out using a restriction enzyme of known sequence specificity, leading to the generation of cleavage sites with known sequence overhangs, suitable ends of the adaptors can be designed to enable hybridization of the adaptor to the cleavage site of the sequence region of interest and subsequent ligation. Ligation also refers to any joining of two nucleic acid molecules that results in a single nucleic acid sequence that can be further modified to obtain the sequence of the nucleic acids in question. Reagents and methods for efficient and rapid ligation of adaptors are commercially available, and are known in the art.

[0046] In some embodiments, the 5' and/or 3' end nucleotide sequences of fragmented nucleic acids are not modified or end-repaired prior to ligation with the adapter oligonucleotides of the present invention. For example, fragmentation by a restriction endonuclease can be used to leave

a predictable overhang, followed by ligation with one or more adapter oligonucleotides comprising an overhang complementary to the predictable overhang on a nucleic acid fragment. In another example, cleavage by an enzyme that leaves a predictable blunt end can be followed by ligation of blunt-ended nucleic acid fragments to adapter oligonucleotides comprising a blunt end. In some embodiments, end repair can be followed by an addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides, such as one or more adenine, one or more thymine, one or more guanine, or one or more cytosine, to produce an overhang. Nucleic acid fragments having an overhang can be joined to one or more adapter oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3' ends of end-repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adapters each having a thymine at a 3' end. In some embodiments, adapter oligonucleotides can be joined to blunt end double-stranded nucleic acid fragments which have been modified by extension of the 3' end with one or more nucleotides followed by 5' phosphorylation. In some cases, extension of the 3' end can be performed with a polymerase such as for example Klenow polymerase or any of the suitable polymerases provided herein, or by use of a terminal deoxynucleotide transferase, in the presence of one or more dNTPs in a suitable buffer containing magnesium. In some embodiments, nucleic acid fragments having blunt ends can be joined to one or more adapters comprising a blunt end. Phosphorylation of 5' ends of nucleic acid fragments can be performed for example with T4 polynucleotide kinase in a suitable buffer containing ATP and magnesium. The fragmented nucleic acid molecules may optionally be treated to dephosphorylate 5' ends or 3' ends, for example, by using enzymes known in the art, such as phosphatases.

[0047] In some embodiments, appending the adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction. In some embodiments, appendage of an adaptor to the nucleic acid fragments comprises ligation. In some embodiments, ligation of the adaptor to the nucleic acid fragments can be following end repair of the nucleic acid fragments. In another embodiment, the ligation of the adaptor to the nucleic acid fragments can be following generation of the nucleic acid fragments without end repair of the nucleic acid fragments. The adaptor can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors in which the adaptor comprises two complementary strands. In some embodiments, the adaptor can be a double stranded DNA adaptor. In some embodiments, the adaptor can be an oligonucleotide of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the adaptor is appended or attached. In some embodiments, the adaptor can be a conventional duplex adaptor, wherein the adaptor comprises sequence well known in the art. In a some embodiments, the adaptor can be appended to the nucleic acid fragments generated by the methods described herein in multiple orientations. In some embodiment, the methods described herein can involve the use of a duplex adaptor comprising double stranded DNA of known sequence that is blunt ended and can bind to the double stranded nucleic acid fragments generated by the methods described herein in one of two orientations. In some embodiments, the adaptor can be ligated to each of the nucleic acid fragments such that each of the nucleic acid fragments comprises the same adaptor. In other words, each of the nucleic acid fragments comprises a common adaptor. In another embodiment, an adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein such that each nucleic acid fragment in the library of nucleic acid

fragments comprises the adaptor ligated to one or both ends. In another embodiment, more than one adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein. The multiple adaptors may occur adjacent to one another, spaced intermittently, or at opposite ends of the nucleic acid fragments. In some embodiments, the adaptor can be ligated or appended to the 5' and/or 3' ends of the nucleic acid fragments generated by the methods described herein. The adaptor can comprise two strands wherein each strand comprises a free 3' hydroxyl group but neither strand comprises a free 5' phosphate. In some embodiments, the free 3' hydroxyl group on each strand of the adaptor can be ligated to a free 5' phosphate present on either end of the nucleic acid fragments of the present invention. In this embodiment, the adaptor comprises a ligation strand and a non-ligation strand whereby the ligation strand can be ligated to the 5' phosphate on either end of the nucleic acid fragment while a nick or gap can be present between the non-ligation strand of the adaptor and the 3' hydroxyl on either end of the nucleic acid fragment. In some embodiments, the nick or gap can be filled in by performing a gap repair reaction. In some embodiments, the gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In some embodiments, the gap repair can be performed using a DNA-dependent DNA polymerase with weak or no strand displacement activity. In some embodiments, the ligation strand of the adaptor can serve as the template for the gap repair or fill-in reaction. The gap repair or fill-in reaction may comprise an extension reaction wherein the ligation strand of the adaptor serves as a template and leads to the generation of nucleic acid fragments with complementary termini or ends. In some embodiments, the gap repair can be performed using Taq DNA polymerase. In some embodiments, the ligation of the first adaptor to the nucleic acid fragments generated by the

methods described herein may not be followed by gap repair. The nucleic acid fragments may comprise the adaptor sequence ligated only at the 5' end of each strand.

[0048] Ligation and, optionally gap repair, of the adaptor to the nucleic acid fragments generates an adaptor-nucleic acid fragment complex. In some embodiments, the adaptor-nucleic acid fragment complex can be denatured. Denaturation can be achieved using any of the methods known in the art including, but not limited to, physical, thermal, and/or chemical denaturation. In some embodiments, denaturation can be achieved using thermal or heat denaturation. In some embodiments, denaturation of the adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising the adaptor sequence at only the 5'end of the nucleic acid fragments. In another embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising adaptor sequence at both the 5'end and 3'end of the nucleic acid fragments.

Methods of Amplification

[0049] The methods, compositions and kits described herein can be useful to generate amplification-ready products directly from a nucleic acid source for downstream applications such as next generation sequencing, as well as generation of libraries with enriched population of sequence regions of interest. In some embodiments, the adapter-nucleic fragment ligated products, e.g., from the ligation of an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, is amplified.

[0050] Methods of amplification are well known in the art. In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal. In

some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR).

[0051] Suitable amplification reactions can be exponential or isothermal and can include any DNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), linear amplification, multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No, 6,251,639), Ribo-SPIA, or a combination thereof. In some cases, the amplification methods for providing the template nucleic acid may be performed under limiting conditions such that only a few rounds of amplification (e.g., 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 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[0052] PCR is an in vitro amplification procedure based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

[0053] LCR uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is

employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[0054] SDA (Westin et a 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3'-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[0055] Some aspects of the invention utilize linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

Downstream Applications

[0056] One aspect of the present invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as in next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. The methods of the present invention can also be used in the analysis of genetic information of selective genomic regions of interest (e.g., analysis of SNPs or other disease markers) as well as genomic regions which may interact with the selective region of interest. The methods of the present invention may further be used in the analysis of copy number variation as well as differential expression.

Sequencing

[0057] In some embodiments, a population of sequencing reads is generated from the amplified adapter-nucleic fragment ligated products. In some embodiments, a sequencing read comprises an index read, which comprises the sequence of an indexing site. In some embodiments, an index read comprises the sequence of an indexing site and the sequence of an identifier site. For example, the indexing site is sequenced with the identifier site. In some embodiments, the index read does not include the sequence of an identifier site. For example, the indexing site is not sequenced with the identifier site. In some embodiments, a sequencing read comprises the target sequence. In some embodiments, a sequencing read comprises a target sequence and an identifier sequence. For example, the target sequence is sequenced with the identifier site. In some embodiments, the target sequence is not sequenced with the identifier site.

[0058] The methods of the present invention are useful for sequencing by the method commercialized by Illumina, as described U.S. Pat Nos. 5,750,341; 6,306,597; and 5,969,119.

[0059] In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g., (A)/(A') or both ends (e.g., (A)/(A') and (C)/(C')). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g., by SPIA or linear PCR).The resulting nucleic acid is then denatured and the single-stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time.

[0060] In some embodiments, the methods of the invention are useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature* (2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711;

US20060024678; US20080213770; and US20080103058. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452; 7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764.

[0061] An example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single head, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

[0062] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g., Soni G V and Meller A. (2007) *Clin Chem*

53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

Data Analysis

[0063] In some embodiments, the sequence reads are used in the analysis of genetic information of selective genomic regions of interest as well as genomic regions which may interact with the selective region of interest. Amplification methods as disclosed herein can be used in the devices, kits, and methods known to the art for genetic analysis, such as, but not limited to those found in U.S. Pat. Nos. 6,449,562, 6,287,766, 7,361,468, 7,414,117, 6,225,109, and 6,110,709.

[0064] In some embodiments, the sequencing reads are used to detect duplicate sequencing reads. In some embodiments, a sequencing read is identified as a duplicate sequencing read when it contains an identifier site and target sequence that is the same as another sequencing read from the same population of sequencing reads.

[0065] In some embodiments, duplicate sequencing reads are differentiated from one another as being true duplicates versus apparent or perceived duplicates. Apparent or perceived duplicates may be identified from sequencing libraries and using conventional measures of duplicate reads (i.e., reads were mapped using bowtie), where all reads with the same start and end nucleic acid coordinates were counted as duplicates. True duplicates may be identified from sequencing

libraries having had an identifier site introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates.

[0066] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acids generated from any dsDNA may have the same start site, as determined from the sequencing reads of the target sequence of the generated nucleic acid fragments. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original dsDNA molecule and, therefore, are no true duplicate reads. The ligation of random sequences onto dsDNA molecules, accompanied by the methods of the invention, allow for the identification of true duplicate reads versus apparent or perceived duplicate reads.

[0067] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acid fragments generated from a genomic DNA (gDNA) molecule that have the same start site, which can be determined from the sequencing reads of the target sequence of the nucleic acid fragments is determined. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original gDNA molecule and, therefore, are not true duplicate reads. In another embodiment, an identifier site is inserted at the adapter insert junction. The sequence of the identifier site is carried through the library amplification step. The identifier site is the first sequence read during the forward read. As the identifier sequence is not logically present adjacent to naturally occurring sequence, this uniquely identifies the DNA fragment. Therefore, by ligating random sequences onto the original gDNA, the methods of the invention identify true duplicate reads.

[0068] In some embodiments, a duplicate sequencing read is detected and analyzed. Duplicate reads can be filtered using 'samtools rmdup', wherein reads with identical external coordinates

are removed, only one read with highest mapping quality is retained. After filtering, a filtered set of deduplicated reads can be used in any downstream analysis. Conversely, this filtering step can be skipped, and downstream analysis can be done using the unfiltered reads, including duplicates.

[0069] In some embodiments, sequencing reads are generated from a number of samples. In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from the samples, in which the nucleic acid fragments from each sample has the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before sequencing reads of the target sequence and/or identifier site are analyzed. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before duplicate sequencing reads are analyzed and/or removed.

[0070] In some embodiments, a method disclosed herein identifies or detects one or more true duplicate(s) with increased accuracy as compared to other methods. For example, in some embodiments, a method disclosed herein identify true duplicates (in contrast to identifying apparent or perceived duplicates) with increased accuracy as compared to other methods. The increased resolution and/or accuracy in identifying one or more true duplicate(s) can provide a considerable contribution to the state of the art in more accurately identifying true duplicates. In some embodiments, a method disclosed herein identifies or detects a true duplicate with

increased efficiency as compared to other methods, such as paired end sequencing. The increase in accuracy, resolution, and/or efficiency in detecting duplicate reads (e.g., true duplicate(s)) can increase confidence in sequencing results, such as for expression and CNV analysis.

Kits

[0071] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for amplification.

[0072] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[0073] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit may be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[0074] A kit may include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

[0075] In some embodiments, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, a kit comprises a composition of the invention, in one or more containers. In some embodiments, the

invention provides kits comprising adapters, primers, and/or other oligonucleotides described herein. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) a forward adapter (e) one or more oligonucleotides comprising reverse adaptor sequence and (f) one or more buffers suitable for one or more of the elements contained in said kit. The adapters, primers, other oligonucleotides, and reagents can be, without limitation, any of those described above. Elements of the kit can further be provided, without limitation, in any of the amounts and/or combinations (such as in the same kit or same container) described above. The kits may further comprise additional agents, such as those described above, for use according to the methods of the invention. For example, the kit can comprise a first forward adaptor that is a partial duplex adaptor as described herein, a second forward adapter, and a nucleic acid modifying enzyme specific for a restriction and/or cleavage site present in the first forward adaptor. The kit elements can be provided in any suitable container, including but not limited to test tubes, vials, flasks, bottles, ampules, syringes, or the like. The agents can be provided in a form that may be directly used in the methods of the invention, or in a form that requires preparation prior to use, such as in the reconstitution of lyophilized agents. Agents may be provided in aliquots for single-use or as stocks from which multiple uses, such as in a number of reaction, may be obtained.

[0076] In some embodiments, the kit comprises a plurality of adaptor oligonucleotides, wherein each of the adaptor oligonucleotides comprises at least one of a plurality of identifier site sequences, wherein each identifier site sequence of the plurality of identifier site sequences differs from every other identifier site sequence in said plurality of identifier site sequences at at least three nucleotide positions, and instructions for using the same. Adapters comprising

different identifier site sequences can be supplied individually or in combination with one or more additional adapters having a different identifier site sequence. In some embodiments, the kit can comprises a plurality of adapter oligonucleotides.

EXAMPLES

Example 1: Identification of Duplicate Sequencing Reads with NuGEN Ovation Target Enrichment Library System

[0077] Sample Description: 100ng of DNA from a human HapMap sample (NA19238) was fragmented to approximately 500 base pair in length by sonication with a Covaris system (Covaris, Inc., Woburn, MA). The resulting DNA was treated with end repair enzyme mix NuGEN R01280 and R01439 (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendation to produce blunt ended DNA fragments.

[0078] Library Generation. Enrichment, and Identifier Site Incorporation: An oligonucleotide with segments, from 5' to 3' of the top strand, 1) an Illumina indexing read priming site such as AGAGCACACGTCTGAACTCCAGTCAC (SEQ ID NO:2), 2) an indexing site, 3) an identifier site having a random 6 base sequence and, 4) a sequence compatible with an Illumina forward sequencing priming site such as TCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO:3) , was annealed to a second oligonucleotide to form a partially double stranded DNA adapter. Five uM of these adapters were ligated onto the end-repaired DNA using Ligase and Ligase reaction buffer from NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendations. Following 30 minutes of incubation at 25°C, the reaction mixture was diluted with water, 0.8X volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated

DNA fragments eluted according to manufacturer's recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95°C then slowly cooling the mixture from 80°C to 60°C by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, MA) according to the manufacturer's protocols. Following extension, the DNA fragments were collected on Agencourt magnetic beads, washed and eluted according to manufacturer's recommendations. These libraries were enriched by 30 cycles of PCR using NuGEN library enrichment primers (Ovation Target Enrichment Library System, NuGEN Technologies, Inc., San Carlos, CA) that also contain the Illumina flow cell sequences (Illumina Inc., San Diego, CA) according to supplier's recommendation.

[0079] The resulting libraries were quantitated by qPCR using a kit provided by KAPA, diluted to 2nM and applied to an Illumina MiSeq DNA Sequencer (Illumina Inc., San Diego, CA). The following series was run: 36 base first read, 14 base second read, and 24 base third read.

[0080] *Data Analysis:* The sequencer output was processed in accordance with manufacturer's recommendation. In order to analyze the data, the indexing read was split into two files. The first file contained the first 8 bases of the indexing reads and is utilized as the library index file for standard library parsing. The other file contains only the random bases and is set aside for further sequence parsing.

[0081] Following our data analysis pipeline of sequence alignment with bowtie aligner (Langmead B. et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009, 10:R2.), duplicate reads were identified by their genomic start positions. At this point, sequencing reads that started at the same genomic position were checked against the random bases file to see if they have the same or different set of random

bases that were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation Target Enrichment targeting probe was used to generate the sequence fragment in question. These sequences, therefore, did not provide unique information about the starting genomic DNA and are considered as one sequencing read for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. **Figure 5** provides the analysis results demonstrating the identification of the duplicate reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

[0082] If the sequences of two libraries are identical, their duplication status is unknown since this could occur by chance in any library. If an identifier sequence is used in combination with the library reads, the status can be determined (duplicates if identical, distinct if different). With the SPET system, one end is common so the probability of two libraries having identical ends increases. These would appear to be duplicate sequences and their true status can be determined by looking at an identifier sequence.

[0083] Over the sampling of all randomly selected reads, the use of identifier sites provided an increased resolution on the presence of true duplicates. When evaluating two million random reads, the apparent duplicates were found to comprise 39% of all reads. However, the true duplicates, identified through the use of an identifier site, were found to comprise only 26% of all reads. The methods employing the use of an identifier site were found to considerably increase the resolution of the true number of duplicates within the pool of reads.

Example 2: Removal of Duplicate Sequencing Reads with 8 Base Identifier Site

[0084] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0085] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites. This identifier site is designed as eight random bases introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain the 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rmdup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 3: Removal of Duplicate Sequencing Reads with Mixture of Random 1-8 Base Identifier Sites

[0086] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the

ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0087] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0088] One to eight random bases are introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain between 1 and 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rmdup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 4: Removal of Duplicate Sequencing Reads with Mixture of 96 Defined 6-Base Identifier Sites

[0089] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0090] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0091] A mixture of 96 defined six-bases sequences are introduced onto the end of the adaptors. Thus, each six-base sequence is an identifier site. Sequence reads from libraries made with such adaptors contain one of the 96 six-base identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rmdup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 5: Identification of Duplicate Sequencing Reads in Determining mRNA Expression Levels

[0092] Sample Description: Total RNA is extracted from tumor and normal adjacent tissue for the purpose of finding differences in expression levels of transcripts between the two sample types. l00ng of each sample is converted into cDNA using the USP primers, reaction buffer, and Reverse Transcriptase provided in NuGEN's Encore Complete Library System (NuGEN Technologies, Inc., San Carlos, CA) according to the supplier's recommendations. This is followed by second strand synthesis, again using materials provided in the kit according to recommendations. Double stranded cDNA was prepared using the Superscript® Double-Stranded cDNA Synthesis Kit from Life Technologies (Carlsbad, CA) according the manufacturer's instructions. DNA was sheared with a Covaris S-series device (Covaris, Inc., Woburn, MA) using the 200 bp sonication protocol provided with the instrument (10% duty cycle, 200 cycles/burst, 5 intensity, 180 seconds). DNA was treated with 1.5 uL 10X Blunting Buffer, 0.5 uL Blunting Enzyme (New England Biolabs, Inc., Ipswich, MA; p/n E1201) and 1.2 uL of 2.5mM of each dNTP mix in a total volume of 15 uL for 30 minutes at 25°C followed by 10 minutes at 70°C.

[0093] *Library Generation*: The DNA fragments were then subjected to end repair using end repair buffers and enzymes provided in NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA).

[0094] Forward adaptor 5'

[0095] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTNNNNNNNN (SEQ ID NO:4),

[0096] Reverse adaptor 1) 5'

[0097] CAAGCAGAAGACGGCATACGAGATTCCCTTGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCTNNNNNNNN (SEQ ID NO:5),

[0098] Reverse adaptor 2) 5'

[0099] CAAGCAGAAGACGGCATACGAGATTGAAGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCNNNNNNNN (SEQ ID NO:6), and

[0100] Common partner 5' NNNNNNNNAGATCGGAAGAGC (SEQ ID NO:7) were all ordered from IDT (Integrated DNA Technologies, Coralville, IA). The reverse adaptors each contain a unique identifier (underlined) that enables libraries made with these adapters to be distinguished. In this case N represents an equimolar mixture of A, C, G, and T. A mixture of 5uM forward, 5uM reverse, and l0uM common in l0mM MgCl2, 50mM Tris pH 8 was heated to 95C for 5 minutes, then cooled to 20C. Adaptor ligation was performed by addition of 4.5 uL water, 3 uL Adaptor mix (prepared above), 6 uL 5X NEBNext Quick Ligation Reaction Buffer and 1.5 uL Quick T4 DNA Ligase (New England Biolabs, Inc., Ipswich, MA; p/n E6056), followed by incubation for 30 minutes at 25°C followed by 10 minutes at 70°C. Ligation products were purified by adding 70uL water and 80 uL of Ampure XP beads (Agencourt Genomics), washing twice with 70% ethanol and eluting with 20 uL of l0mM Tris pH 8.0.

Library products were amplified in a 50uL PCR containing 0.5uM each of primer (5' AATGATACGGCGACCACCGA (SEQ ID NO:8), and 5' CAAGCAGAAGACGGCATACGA (SEQ ID NO:9), 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM MgCl2, 0.2mM each dNTP, and 1 unit Taq polymerase. The reaction was cycled 15 times under the conditions 95C for 15 seconds, 60C for 1 minute. PCR products were purified with 1 volume of Ampure XP beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) as described above. The library was analyzed by HS DNA Bioanalyzer (Agilent Technologies, Santa Clara, CA) and quantitated with the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA; p/n KK4835) according to the supplied instructions. The resulting libraries are combined and are compatible with standard TruSeq single end or paired Illumina sequencing protocols for GAIIx, MiSeq, or Hi Seq sequencing instruments (Illumina Inc., San Diego, CA). The following series is run: 50 base first read, 6 base second read. A third read is not required for counting purposes or duplication analysis.

[0101] Data Analysis: The sequencer output was processed in accordance with manufacturer's recommendation. The 6 bases of the index read is used for standard library parsing, separating the data files from the two sample types. The first 50 bases of the target sequence read are compared to each other. Any read that has identical sequence to any other read is identified as a duplicate and removed from the population, thus, only a single copy is retained within the file. Once the duplicate reads have been removed, 8 bases are trimmed from each read. The trimmed reads are aligned to a reference genome. Differential expression is then determined by comparing FPKM (fragments per kilobase per million reads) values between libraries utilizing scripts such as cufflinks or cuffdiff (Trapnell et al. 2010, Nature Biotechnology, 28, 511-515; Trapnell et al. 2013, Nature Biotechnology, 31, 46-53).

Example 6: Sequencing of the Identifier Site With the Target Sequence in Paired End Reads

[0102] The Ovation Library System for Low Complexity Samples (NuGEN Technologies, Inc., San Carlos, CA) was used to generate four libraries, each from a single amplicon, following manufacturer's protocol. Purified libraries were mixed and sequenced as a multiplex on the Illumina MiSeq (Illumina Inc., San Diego, CA) to produce 125nt forward, 8nt index 1, 8nt index 2, and 25nt reverse reads. Because all amplicon reads start and end at the same sequence coordinates, the traditional method of marking library PCR duplicates (marking reads that start and end at the same genomic coordinates as duplicates) cannot be used. Instead, the 0-8nt of random sequence contained in the adaptors ligated to the amplicon were treated as an identifier sequence and used to mark duplicates. Any paired end reads that shared the same length and sequence of these random bases with any other paired end read was called a duplicate. The table below shows the results of this duplicate marking.

Table 1.

Library	Reads	Duplicate Reads	Reads from independent molecules
1	454249	376797	77452
2	439367	364001	75366
3	317760	253057	64703
4	476572	398380	78192

[0103] Table 1 demonstrates the accuracy of the method used to differentiate between duplicate reads and reads from truly independent molecules. The population of reads from independent molecules, depicted in the final column of Table 1, represent sequences from independent amplicon molecules used in generating the libraries.

Example 7: Sequencing of the Identifier Site With the Target Sequence in Reduced Representation Bisulfite (RRBS) Libraries

[0104] Reduced representation bisulfite (RRBS) libraries of the human genome are generated through the complete restriction enzyme digestion of 100ng input sample, followed by selection for short fragments. The resulting pool of fragments are ligated to adaptor sequences comprising an indexing site and an identifier site. The identifier sites comprise either 6 or 8 random nucleotides. The sequences are then sequenced to identify the identifier sites, thus revealing the true number of duplicates in the pool. In the absence of identifier sites, the number of apparent or perceived duplicates are greater than the number of true duplicates. The inclusion of an identifier site results in the identification of the number of true duplicates, as compared to the larger number of apparent or perceived duplicates.

[0105] Unless defined otherwise, all technical and scientific terms herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials, similar or equivalent to those described herein, can be used in the practice or testing of the present invention, the preferred methods and materials are described herein. All publications, patents, and patent publications cited are incorporated by reference herein in their entirety for all purposes.

[0106] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

[0107] All references, articles, publications, patents, patent publications, and patent applications cited herein are incorporated by reference in their entireties for all purposes. However, mention of any reference, article, publication, patent, patent publication, and

patent application cited herein is not, and should not be taken as an acknowledgment or any form of suggestion that they constitute valid prior art or form part of the common general knowledge in any country in the world.

[0108] While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth and as follows in the scope of the appended claims.

IDT Exhibit 1019.0173

Amendments to the Specification

Please replace the specification of the subject application, i.e., pages 1-40, with the substitute specification filed herewith.

IDT Exhibit 1019.0174

Sequence Listing

Please enter the Sequence Listing in computer readable form (CRF), filed herewith, which is contained in the text file named: NUGN-001-03US_SeqList_ST25.txt.

IDT Exhibit 1019.0175

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 16/743,604

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			N/A	300
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A			N/A	660
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	760
TOTAL CLAIMS (37 CFR 1.16(i))	11 minus 20 =	*			OR	x 100 =	0.00
INDEPENDENT CLAIMS (37 CFR 1.16(h))	1 minus 3 =	*				x 460 =	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			TOTAL	1720

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	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

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

UNITED STATES PATENT AND TRADEMARK OFFICE

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

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/743,604	01/15/2020		1820	NUGN-001/03US 34079/516	11	1

CONFIRMATION NO. 3249

UPDATED FILING RECEIPT

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

OC000000118374777

Date Mailed: 04/22/2020

Receipt is acknowledged of this non-provisional utility 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 FIRST INVENTOR, 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 corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction 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 provided that the request is grantable.

Inventor(s)

Douglas A. Amorese, Los Altos, CA;
Jonathan Scolnick, San Francisco, CA;
Benjamin G. Schroeder, San Mateo, CA;

Applicant(s)

NUGEN TECHNOLOGIES, INC., Redwood City, CA;

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 15/406,002 01/13/2017 PAT 10570448
which is a DIV of 14/540,917 11/13/2014 PAT 9546399
which claims benefit of 61/903,826 11/13/2013

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

IDT Exhibit 1019.0177

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: 02/14/2020
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/743,604**
Projected Publication Date: 07/30/2020
Non-Publication Request: No
Early Publication Request: No
Title

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

Preliminary Class

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

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

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

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

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

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

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

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

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

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

NOT GRANTED

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

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 December 2008 (11.12.2008)

PCT

(10) International Publication Number
WO 2008/150432 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2008/006822

(22) International Filing Date: 29 May 2008 (29.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/941,381 1 June 2007 (01.06.2007) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/941,381 (CIP)
Filed on 1 June 2007 (01.06.2007)

(71) Applicant *(for all designated States except US)*: **454 LIFE SCIENCES CORPORATION** [US/US]; 20 Commercial Street, Branford, CT 06405 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BRAVERMAN, Michael, S.** [US/US]; 375 Stevenson Road, New Haven, CT 06515 (US). **SIMONS, Jan, Fredrik** [FI/US]; 2515 Poppy Drive, Burlingame, CA 94010 (US). **SRINIVASAN, Maithreyan** [US/US]; 801 Church Street #1398, Mountain View, CA 94041 (US). **TURENCHALK, Gregory, S.** [US/US]; 1448 Quinnipiac Avenue, New Haven, CT 06513 (US).

(74) Agents: **IWAMOTO, Michelle, A.** et al.; Mintz, Levin, Cohn, Ferris, glovsky And Popeo P.C., Chrysler Center, 666 Third Avenue, New York, NY 10017 (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, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US (patent), UZ, VC, VN, ZA, ZM, ZW.

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

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

WO 2008/150432 A1

(54) Title: SYSTEM AND METH0D FOR IDENTIFICATION OF INDIVIDUAL SAMPLES FROM A MULTIPLEX MIXTURE

(57) Abstract: An embodiment of an identifier element for identifying an origin of a template nucleic acid molecule is described that comprises a nucleic acid element comprising a sequence composition that enables detection of an introduced error in sequence data generated from the nucleic acid element and correction of the introduced error, where the nucleic acid element is constructed to couple with the end of a template nucleic acid molecule and identifies an origin of the template nucleic acid molecule.

IDT Exhibit 1019.0180

SYSTEM AND METHOD FOR IDENTIFICATION OF INDIVIDUAL SAMPLES FROM A MULTIPLEX MIXTURE

FIELD OF THE INVENTION

The present invention relates to the fields of molecular biology and bioinformatics. More specifically, the invention relates to associating a unique identifier (UID) element, which is sometimes also referred to as a multiplex identifier (MID), with one or more nucleic acid elements derived from a specific sample, combining the associated elements from the sample with associated elements from one or more other samples into a multiplex mixture of said samples, and identifying each identifier and its associated sample from data generated by what are generally referred to as "Sequencing" techniques.

BACKGROUND OF THE INVENTION

There are a number of "sequencing" techniques known in the art amenable for use with the presently described invention such as, for instance, techniques based upon what are referred to as Sanger sequencing methods commonly known to those of ordinary skill in the art that employ termination and size separation techniques. Other classes of powerful high throughput sequencing techniques for determining the identity or sequence composition of one or more nucleotides in a nucleic acid sample include what are referred to as "Sequencing-by-synthesis" techniques (SBS), "Sequencing-by-Hybridization" (SBH), or "Sequencing-by-Ligation" (SBL) techniques. Of these, SBS methods provide many desirable advantages over previously employed sequencing methods that include, but are not limited to the massively parallel generation of a large volume of high quality sequence information at a low cost relative to previous techniques. The term "massively parallel" as used herein generally refers to the simultaneous generation of sequence information from many different template molecules in parallel where the individual template molecule or population of substantially identical template molecules are separated or compartmentalized and simultaneously exposed to sequencing processes which may include a iterative series of reactions thereby producing an independent sequence read representing the nucleic acid composition of each template molecule. In other words, the advantage includes the ability to simultaneously sequence multiple nucleic acid elements associated with many different samples or different nucleic acid elements existing within a sample.

Typical embodiments of SBS methods comprise the stepwise synthesis of a single strand of polynucleotide molecule complementary to a template nucleic acid molecule whose

IDT Exhibit 1019.0181

nucleotide sequence composition is to be determined. For example, SBS techniques typically operate by adding a single nucleic acid (also referred to as a nucleotide) species to a nascent polynucleotide molecule complementary to a nucleic acid species of a template molecule at a corresponding sequence position. The addition of the nucleic acid species to the nascent molecule is generally detected using a variety of methods known in the art that include, but are not limited to what are referred to as pyrosequencing or fluorescent detection methods such as those that employ reversible terminators or energy transfer labels including fluorescent resonant energy transfer dyes (FRET). Typically, the process is iterative until a complete (i.e. all sequence positions are represented) or desired sequence length complementary to the template is synthesized.

Further, as described above many embodiments of SBS are enabled to perform sequencing operations in a massively parallel manner. For example, some embodiments of SBS methods are performed using instrumentation that automates one or more steps or operation associated with the preparation and/or sequencing methods. Some instruments employ elements such as plates with wells or other type of microreactor configuration that provide the ability to perform reactions in each of the wells or microreactors simultaneously. Additional examples of SBS techniques as well as systems and methods for massively parallel sequencing are described in US Patent Nos. 6,274,320; 6,258,568; 6,210,891, 7,211,390; 7,244,559; 7,264,929; 7,335,762; and 7,323,305 each of which is hereby incorporated by reference herein in its entirety for all purposes; and US Patent Application Serial No. 11/195,254, which is hereby incorporated by reference herein in its entirety for all purposes.

It may also be desirable in some embodiments of SBS, to generate many substantially identical copies of each template nucleic acid element that for instance, provides a stronger signal when one or more nucleotide species is incorporated in each nascent molecule in a population comprising the copies of a template nucleic acid molecule. There are many techniques known in the art for generating copies of nucleic acid molecules such as, for instance, amplification using what are referred to as bacterial vectors, "Rolling Circle" amplification (described in US Patent Nos. 6,274,320 and 7,211,390, incorporated by reference above), isothermal amplification techniques, and Polymerase Chain Reaction (PCR) methods, each of the techniques are applicable for use with the presently described invention. One PCR technique that is particularly amenable to high throughput applications include what are referred to as emulsion PCR methods.

IDT Exhibit 1019.0182

Typical embodiments of emulsion PCR methods include creating stable emulsion of two immiscible substances and are resistant to blending together where one substance is dispersed within a second substance. The emulsions may include droplets suspended within another fluid and are sometimes also referred to as compartments, microcapsules, microreactors, microenvironments, or other name commonly used in the related art. The droplets may range in size depending on the composition of the emulsion components and formation technique employed. The described emulsions create the microenvironments within which chemical reactions, such as PCR, may be performed. For example, template nucleic acids and all reagents necessary to perform a desired PCR reaction may be encapsulated and chemically isolated in the droplets of an emulsion. Thermo cycling operations typical of PCR methods may be executed using the droplets to amplify an encapsulated nucleic acid template resulting in the generation of a population comprising many substantially identical copies of the template nucleic acid. Also in the present example, some or all of the described droplets may further encapsulate a solid substrate such as a bead for attachment of nucleic acids, reagents, labels, or other molecules of interest.

Embodiments of an emulsion useful with the presently described invention may include a very high density of droplets or microcapsules enabling the described chemical reactions to be performed in a massively parallel way. Additional examples of emulsions and their uses for sequencing applications are described in US Patent Application Serial Nos. 10/861,930; 10/866,392; 10/767,899; 11/045,678 each of which are hereby incorporated by reference herein in its entirety for all purposes.

Those of ordinary skill in the related art will appreciate that advantages provided by the massively parallel nature of the amplification and sequencing methods described herein may be particularly to amenable for processing what may be referred to as a "Multiplex" sample. For example, a multiplex composition may include representatives from multiple samples such as samples from multiple individuals. It may be desirable in many applications to combine multiple samples into a single multiplexed sample that may be processed in one operation as opposed to processing each sample separately. Thus the result may typically include a substantial savings in reagent, labor, and instrument usage and cost as well as a significant savings in processing time invested. The described advantages of multiplex processing become more pronounced as the numbers of individual samples increase. Further, multiplex processing has application in research as well as diagnostic contexts. For example, it may be desirable in many applications to employ a single multiplexed sample in an

IDT Exhibit 1019.0183

amplification reaction and subsequently processing the amplified multiplex composition in a single sequencing run.

One problem associated with processing a multiplex composition then becomes identifying the association between each sample of origin and the sequence data generated from a template molecule derived from said sample. A solution to this problem includes associating an identifier such as a nucleic acid sequence that specifically identifies the association of each template molecule with its sample of origin. An advantage of this solution is that the sequence information of the associated nucleic acid sequence is embedded in the sequence data generated from the template molecule and may be bioinformatically analyzed to associate the sequence data with its sample of origin.

Previous studies have described associating nucleic acid sequence identifiers with 5' primers coupled with target sequences for multiplex processing. One such study is that of Binladen et al. (Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C (2007) The use of coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by Parallel 454 Sequencing. PLoS ONE 2(2): e197.doi:10.1371/journal.pone.0000197 (published online February 14, 2007, which is hereby incorporated by reference herein in its entirety for all purposes). As mentioned above, Binladen et al. describe associating short sequence identifiers with target sequences to be processed in a multiplex sample producing sequence data that is subsequently bioinformatically analyzed to associate the short identifiers with their sample of origin. However, there are limitations to simply attaching a nucleic acid identifier of generic sequence composition to a template molecule and identifying the sequence of said identifier in the generated sequence data. Of primary concern is the introduction of error into the sequence data from various mechanisms. Such mechanisms typically work in combination with each other and are generally not individually identifiable from the sequence data. Thus because of introduced error, an end user may not be able to identify the association between the sequence data with its sample of origin, or possibly worse fail to identify that an error has occurred and mis-assign sequence data to a sample of origin that is incorrect.

There are two important sources of error introduction to consider, although other sources may also exist. First is error introduced by the sequencing operation that may in some cases be referred to a "flow error". For example, flow error may include polymerase errors that include incorporation of an incorrect nucleotide species by a polymerase enzyme. A sequencing operation may also introduce what may be referred to as phasic synchrony error that include what are referred to as "carry forward" and "incomplete extension" (the

IDT Exhibit 1019.0184

combination of phasic synchrony error is sometimes referred to as CAFIE error). Phasic synchrony error and methods of correction are further described in PCT Application Serial No. US2007/004187, titled "System and Method for Correcting Primer Extension Errors in Nucleic Acid Sequence Data", filed February 15, 2007 which is hereby incorporated by reference herein in its entirety for all purposes.

Second is error introduced from processes that are independent of the sequencing operations such as primer synthesis or amplification error. For example, oligonucleotide primers synthesized for PCR may include one or more UID elements of the presently described invention, where error may be introduced in the synthesis of the primer/UID element that is then employed as a sequencing template. High fidelity sequencing of the UID element faithfully reproduces the synthesized error in sequence data. Also in the present example, polymerase enzymes commonly employed in PCR methods are known for having a measure of replication error, where for instance an error in replication may be introduced by the polymerase in 1 of every 10,000; 100,000; or 1,000,000 bases amplified.

Therefore, it is significantly advantageous to employ unique identifiers that are 1) resistant to error introduction; 2) enable detection of introduced error; and 3) enable correction of introduced error. The presently described invention addresses these problems and provides systems and methods for associating unique identifiers that provide better recognition and identification characteristics resulting in improved data quality and experimental efficiency.

SUMMARY OF THE INVENTION

Embodiments of the invention relate to the determination of the sequence of nucleic acids. More particularly, embodiments of the invention relate to methods and systems for correcting errors in data obtained during the sequencing of nucleic acids and associating the nucleic acids with their origin.

An embodiment of an identifier element for identifying an origin of a template nucleic acid molecule is described that comprises a nucleic acid element comprising a sequence composition that enables detection of an introduced error in sequence data generated from the nucleic acid element and correction of the introduced error, where the nucleic acid element is constructed to couple with the end of a template nucleic acid molecule and identifies an origin of the template nucleic acid molecule.

Also, an embodiment of a method for identifying an origin of a template nucleic acid molecule is described that comprises the steps of identifying a first identifier sequence from

IDT Exhibit 1019.0185

sequence data generated from a template nucleic acid molecule; detecting an introduced error in the first identifier sequence; correcting the introduced error in the first identifier sequence; associating the corrected first identifier sequence with a first identifier element coupled to the template molecule; and identifying an origin of the template molecule using the association of the corrected first identifier sequence with the first identifier element.

In some implementations, the method further comprises the steps of identifying a second identifier sequence from the sequence data generated from the template nucleic acid molecule; detecting an introduced error in the second identifier sequence; correcting the introduced error in the second identifier sequence; associating the corrected second identifier sequence with a second identifier element coupled with the template nucleic acid molecule; and identifying an origin of the template nucleic acid molecule using the association of the corrected second identifier sequence with the second identifier element combinatorially with the association of the corrected first identifier sequence with the first identifier element.

Further, an embodiment of a kit for identifying an origin of a template nucleic acid molecule is described that comprises a set of nucleic acid elements each comprising a distinctive sequence composition that enables detection of an introduced error in sequence data generated from each nucleic acid element and correction of the introduced error, wherein each of the nucleic acid elements is constructed to couple with the end of a template nucleic acid molecule and identifies the origin of the template nucleic acid molecule.

In addition, an embodiment of a computer comprising executable code stored in system memory is described where the executable code performs a method for identifying an origin of a template nucleic acid molecule comprising the steps of identifying an identifier sequence from sequence data generated from a template nucleic acid molecule; detecting an introduced error in the identifier sequence; correcting the introduced error in the identifier sequence; associating the corrected identifier sequence with an identifier element coupled with the template molecule; and identifying an origin of the template molecule using the association of the corrected identifier sequence with the identifier element.

The above embodiments and implementations are not necessarily inclusive or exclusive of each other and may be combined in any manner that is non-conflicting and otherwise possible, whether they be presented in association with a same, or a different, embodiment or implementation. The description of one embodiment or implementation is not intended to be limiting with respect to other embodiments and/or implementations. Also, any one or more function, step, operation, or technique described elsewhere in this specification may, in alternative implementations, be combined with any one or more

IDT Exhibit 1019.0186

function, step, operation, or technique described in the summary. Thus, the above embodiment and implementations are illustrative rather than limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

The above and further features will be more clearly appreciated from the following detailed description when taken in conjunction with the accompanying drawings. In the drawings, like reference numerals indicate like structures, elements, or method steps and the leftmost digit of a reference numeral indicates the number of the figure in which the references element first appears (for example, element 160 appears first in Figure 1). All of these conventions, however, are intended to be typical or illustrative, rather than limiting.

Figure 1 is a functional block diagram of one embodiment of a sequencing instrument and computer system amenable for use with the presently described invention;

Figure 2A is a simplified graphical representation of one embodiment of an adaptor element amenable for use with genomic libraries comprising a UID component;

Figure 2B is a simplified graphical representation of one embodiment of an adaptor element amenable for use with amplicons comprising a UID component; and

Figure 3 is a simplified graphical representation of one embodiment of computed error balls representing compatibility of UID elements of different sequence composition.

DETAILED DESCRIPTION OF THE INVENTION

As will be described in greater detail below, embodiments of the presently described invention include systems and methods for associating a unique identifier hereafter referred to as a UID element with one or more nucleic acid molecules from a sample. The UID elements are resistant to introduced error in sequence data, and enable detection and correction of error. Further, the invention includes combining or pooling those UID associated nucleic acid molecules with similarly UID associated (sometimes also referred to as "labeled") nucleic acid molecules from one or more other samples, and sequencing each nucleic acid molecule in the pooled sample to generate sequence data for each nucleic acid. The presently described invention further includes systems and methods for designing the sequence composition for each UID element and analyzing the sequence data of each nucleic acid to identify an embedded UID sequence code and associating said code with the sample identity.

a. General

The terms "flowgram" and "pyrogram" may be used interchangeably herein and generally refer to a graphical representation of sequence data generated by SBS methods.

Further, the term "read" or "sequence read" as used herein generally refers to the entire sequence data obtained from a single nucleic acid template molecule or a population of a plurality of substantially identical copies of the template nucleic acid molecule.

The terms "run" or "sequencing run" as used herein generally refer to a series of sequencing reactions performed in a sequencing operation of one or more template nucleic acid molecule.

The term "flow" as used herein generally refers to a serial or iterative cycle of addition of solution to an environment comprising a template nucleic acid molecule, where the solution may include a nucleotide species for addition to a nascent molecule or other reagent such as buffers or enzymes that may be employed to reduce carryover or noise effects from previous flow cycles of nucleotide species.

The term "flow cycle" as used herein generally refers to a sequential series of flows where a nucleotide species is flowed once during the cycle (i.e. a flow cycle may include a sequential addition in the order of T, A, C, G nucleotide species, although other sequence combinations are also considered part of the definition). Typically the flow cycle is a repeating cycle having the same sequence of flows from cycle to cycle.

The term "read length" as used herein generally refers to an upper limit of the length of a template molecule that may be reliably sequenced. There are numerous factors that contribute to the read length of a system and/or process including, but not limited to the degree of GC content in a template nucleic acid molecule.

A "nascent molecule" generally refers to a DNA strand which is being extended by the template-dependent DNA polymerase by incorporation of nucleotide species which are complementary to the corresponding nucleotide species in the template molecule.

The terms "template nucleic acid", "template molecule", "target nucleic acid", or "target molecule" generally refer to a nucleic acid molecule that is the subject of a sequencing reaction from which sequence data or information is generated.

The term "nucleotide species" as used herein generally refers to the identity of a nucleic acid monomer including purines (Adenine, Guanine) and pyrimidines (Cytosine, Uracil, Thymine) typically incorporated into a nascent nucleic acid molecule.

The term "monomer repeat" or "homopolymers" as used herein generally refers to two or more sequence positions comprising the same nucleotide species (i.e. a repeated nucleotide species).

IDT Exhibit 1019.0188

The term "homogeneous extension", as used herein, generally refers to the relationship or phase of an extension reaction where each member of a population of substantially identical template molecules is homogenously performing the same extension step in the reaction.

The term "completion efficiency" as used herein generally refers to the percentage of nascent molecules that are properly extended during a given flow.

The term "incomplete extension rate" as used herein generally refers to the ratio of the number of nascent molecules that fail to be properly extended over the number of all nascent molecules.

The term "genomic library" or "shotgun library" as used herein generally refers to a collection of molecules derived from and/or representing an entire genome (i.e. all regions of a genome) of an organism or individual.

The term "amplicon" as used herein generally refers to selected amplification products such as those produced from Polymerase Chain Reaction or Ligase Chain Reaction techniques.

The term "keypass" or "keypass mapping" as used herein generally refers to a nucleic acid "key element" associated with a template nucleic acid molecule in a known location (i.e. typically included in a ligated adaptor element) comprising known sequence composition that is employed as a quality control reference for sequence data generated from template molecules. The sequence data passes the quality control if it includes the known sequence composition associated with a Key element in the correct location.

The term "blunt end" or "blunt ended" as used herein generally refers to a linear double stranded nucleic acid molecule having an end that terminates with a pair of complementary nucleotide base species, where a pair of blunt ends are always compatible for ligation to each other.

Some exemplary embodiments of systems and methods associated with sample preparation and processing, generation of sequence data, and analysis of sequence data are generally described below, some or all of which are amenable for use with embodiments of the presently described invention. In particular the exemplary embodiments of systems and methods for preparation of template nucleic acid molecules, amplification of template molecules, generating target specific amplicons and/or genomic libraries, sequencing methods and instrumentation, and computer systems are described.

In typical embodiments, the nucleic acid molecules derived from an experimental or diagnostic sample must be prepared and processed from its raw form into template molecules

IDT Exhibit 1019.0189

amenable for high throughput sequencing. The processing methods may vary from application to application resulting in template molecules comprising various characteristics. For example, in some embodiments of high throughput sequencing it is preferable to generate template molecules with a sequence or read length that is at least the length a particular sequencing method can accurately produce sequence data for. In the present example, the length may include a range of about 25-30 base pairs, about 30-50 base pairs, about 50-100 base pairs, about 100-200 base pairs, about 200-300 base pairs, or about 350-500 base pairs, or other length amenable for a particular sequencing application. In some embodiments, nucleic acids from a sample, such as a genomic sample, are fragmented using a number of methods known to those of ordinary skill in the art. In preferred embodiments, methods that randomly fragment (i.e. do not select for specific sequences or regions) nucleic acids are employed that include what is referred to as nebulization or sonication. It will however, be appreciated that other methods of fragmentation such as digestion using restriction endonucleases may be employed for fragmentation purposes. Also in the present example, some processing methods may employ size selection methods known in the art to selectively isolate nucleic acid fragments of the desired length.

Also, it is preferable in some embodiments to associate additional functional elements with each template nucleic acid molecule. The elements may be employed for a variety of functions including, but not limited to, primer sequences for amplification and/or sequencing methods, quality control elements, unique identifiers that encode various associations such as with a sample of origin or patient, or other functional element. For example, some embodiments may associate priming sequence elements or regions comprising complementary sequence composition to primer sequences employed for amplification and/or sequencing. Further, the same elements may be employed for what may be referred to as "strand selection" and immobilization of nucleic acid molecules to a solid phase substrate. In the present example, two sets of priming sequence regions (hereafter referred to as priming sequence A, and priming sequence B) may be employed for strand selection where only single strands having one copy of priming sequence A and one copy of priming sequence B is selected and included as the prepared sample. The same priming sequence regions may be employed in methods for amplification and immobilization where, for instance priming sequence B may be immobilized upon a solid substrate and amplified products are extended therefrom.

Additional examples of sample processing for fragmentation, strand selection, and addition of functional elements and adaptors are described in U.S. Patent Application Serial

IDT Exhibit 1019.0190

No. 10/767,894, titled "Method for preparing single-stranded DNA libraries", filed January 28, 2004; and U.S. Provisional Application Serial No. 60/941,381, titled "System and Method for Identification of Individual Samples from a Multiplex Mixture", filed June 1, 2007, each of which is hereby incorporated by reference herein in its entirety for all purposes.

Various examples of systems and methods for performing amplification of template nucleic acid molecules to generate populations of substantially identical copies are described. It will be apparent to those of ordinary skill that it is desirable in some embodiments of SBS to generate many copies of each nucleic acid element to generate a stronger signal when one or more nucleotide species is incorporated into each nascent molecule associated with a copy of the template molecule. There are many techniques known in the art for generating copies of nucleic acid molecules such as, for instance, amplification using what are referred to as bacterial vectors, "Rolling Circle" amplification (described in US Patent Nos. 6,274,320 and 7,211,390, incorporated by reference above) and Polymerase Chain Reaction (PCR) methods, each of the techniques are applicable for use with the presently described invention. One PCR technique that is particularly amenable to high throughput applications include what are referred to as emulsion PCR methods (also referred to as emPCR™ methods).

Typical embodiments of emulsion PCR methods include creating a stable emulsion of two immiscible substances creating aqueous droplets within which reactions may occur. In particular, the aqueous droplets of an emulsion amenable for use in PCR methods may include a first fluid such as a water based fluid suspended or dispersed in what may be referred to as a discontinuous phase within another fluid such as an oil based fluid. Further, some emulsion embodiments may employ surfactants that act to stabilize the emulsion that may be particularly useful for specific processing methods such as PCR. Some embodiments of surfactant may include non-ionic surfactants such as sorbitan monooleate (also referred to as Span™ 80), polyoxyethylenesorbitsan monooleate (also referred to as Tween™ 80), or in some preferred embodiments dimethicone copolyol (also referred to as Abil® EM90), polysiloxane, polyalkyl polyether copolymer, polyglycerol esters, poloxamers, and PVP/hexadecane copolymers (also referred to as Unimer U-151), or in more preferred embodiments a high molecular weight silicone polyether in cyclopentasiloxane (also referred to as DC 5225C available from Dow Corning).

The droplets of an emulsion may also be referred to as compartments, microcapsules, microreactors, microenvironments, or other name commonly used in the related art. The aqueous droplets may range in size depending on the composition of the emulsion components or composition, contents contained therein, and formation technique employed.

The described emulsions create the microenvironments within which chemical reactions, such as PCR, may be performed. For example, template nucleic acids and all reagents necessary to perform a desired PCR reaction may be encapsulated and chemically isolated in the droplets of an emulsion. Additional surfactants or other stabilizing agent may be employed in some embodiments to promote additional stability of the droplets as described above. Thermocycling operations typical of PCR methods may be executed using the droplets to amplify an encapsulated nucleic acid template resulting in the generation of a population comprising many substantially identical copies of the template nucleic acid. In some embodiments, the population within the droplet may be referred to as a "clonally isolated", "compartmentalized", "sequestered", "encapsulated", or "localized" population. Also in the present example, some or all of the described droplets may further encapsulate a solid substrate such as a bead for attachment of template or other type of nucleic acids, reagents, labels, or other molecules of interest.

Embodiments of an emulsion useful with the presently described invention may include a very high density of droplets or microcapsules enabling the described chemical reactions to be performed in a massively parallel way. Additional examples of emulsions employed for amplification and their uses for sequencing applications are described in US Patent Application Serial Nos. 10/861,930; 10/866,392; 10/767,899; 11/045,678 each of which are hereby incorporated by reference herein in its entirety for all purposes.

Also, an exemplary embodiment for generating target specific amplicons for sequencing is described that includes using sets of nucleic acid primers to amplify a selected target region or regions from a sample comprising the target nucleic acid. Further, the sample may include a population of nucleic acid molecules that are known or suspected to contain sequence variants and the primers may be employed to amplify and provide insight into the distribution of sequence variants in the sample.

For example a method for identifying a sequence variant by specific amplification and sequencing of multiple alleles in a nucleic acid sample may be performed. The nucleic acid is first subjected to amplification by a pair of PCR primers designed to amplify a region surrounding the region of interest or segment common to the nucleic acid population. Each of the products of the PCR reaction (amplicons) is subsequently further amplified individually in separate reaction vessels such as an emulsion based vessel described above. The resulting amplicons (referred to herein as second amplicons), each derived from one member of the first population of amplicons, are sequenced and the collection of sequences, from different emulsion PCR amplicons, are used to determine an allelic frequency.

IDT Exhibit 1019.0192

Some advantages of the described target specific amplification and sequencing methods include a higher level of sensitivity than previously achieved. Further, embodiments that employ high throughput sequencing instrumentation such as for instance embodiments that employ what is referred to as a PicoTiterPlate® array of wells provided by 454 Life Sciences Corporation, the described methods can be employed to sequence over 100,000 or over 300,000 different copies of an allele per run or experiment. Also, the described methods provide a sensitivity of detection of low abundance alleles which may represent 1% or less of the allelic variants. Another advantage of the methods includes generating data comprising the sequence of the analyzed region. Importantly, it is not necessary to have prior knowledge of the sequence of the locus being analyzed.

Additional examples of target specific amplicons for sequencing are described in U.S. Patent Application Serial No. 11/104,781, titled "Methods for determining sequence variants using ultra-deep sequencing", filed April 12, 2005, which is hereby incorporated by reference herein in its entirety for all purposes.

Further, embodiments of sequencing may include Sanger type techniques, what is referred to as polony sequencing techniques, nanopore and other single molecule detection techniques, or reversible terminator techniques. As described above a preferred technique may include Sequencing by Synthesis methods. For example, some SBS embodiments sequence populations of substantially identical copies of a nucleic acid template and typically employ one or more oligonucleotide primers designed to anneal to a predetermined, complementary position of the sample template molecule or one or more adaptors attached to the template molecule. The primer/template complex is presented with a nucleotide species in the presence of a nucleic acid polymerase enzyme. If the nucleotide species is complementary to the nucleic acid species corresponding to a sequence position on the sample template molecule that is directly adjacent to the 3' end of the oligonucleotide primer, then the polymerase will extend the primer with the nucleotide species. Alternatively, in some embodiments the primer/template complex is presented with a plurality of nucleotide species of interest (typically A, G, C, and T) at once, and the nucleotide species that is complementary at the corresponding sequence position on the sample template molecule directly adjacent to the 3' end of the oligonucleotide primer is incorporated. In either of the described embodiments, the nucleotide species may be chemically blocked (such as at the 3'-O position) to prevent further extension, and need to be deblocked prior to the next round of synthesis. It will also be appreciated that the process of adding a nucleotide species to the

IDT Exhibit 1019.0193

end of a nascent molecule is substantially the same as that described above for addition to the end of a primer.

As described above, incorporation of the nucleotide species can be detected by a variety of methods known in the art, e.g. by detecting the release of pyrophosphate (PPi) (examples described in US Patent Nos. 6,210,891; 6,258,568; and 6,828,100, each of which is hereby incorporated by reference herein in its entirety for all purposes), or via detectable labels bound to the nucleotides. Some examples of detectable labels include but are not limited to mass tags and fluorescent or chemiluminescent labels. In typical embodiments, unincorporated nucleotides are removed, for example by washing. Further, in some embodiments the unincorporated nucleotides may be subjected to enzymatic degradation such as, for instance, degradation using the apyrase enzyme as described in U.S. Provisional Patent Application Serial No. 60/946,743, titled System and Method For Adaptive Reagent Control in Nucleic Acid Sequencing, filed June 28, 2007, which is hereby incorporated by reference herein in its entirety for all purposes. In the embodiments where detectable labels are used, they will typically have to be inactivated (e.g. by chemical cleavage or photobleaching) prior to the following cycle of synthesis. The next sequence position in the template/polymerase complex can then be queried with another nucleotide species, or a plurality of nucleotide species of interest, as described above. Repeated cycles of nucleotide addition, extension, signal acquisition, and washing result in a determination of the nucleotide sequence of the template strand. Continuing with the present example, a large number or population of substantially identical template molecules (e.g. 10^3, 10^4, 10^5, 10^6 or 10^7 molecules) are typically analyzed simultaneously in any one sequencing reaction, in order to achieve a signal which is strong enough for reliable detection.

In addition, it may be advantageous in some embodiments to improve the read length capabilities and qualities of a sequencing process by employing what may be referred to as a "paired-end" sequencing strategy. For example, some embodiments of sequencing method have limitations on the total length of molecule from which a high quality and reliable read may be generated. In other words, the total number of sequence positions for a reliable read length may not exceed 25, 50, 100, or 150 bases depending on the sequencing embodiment employed. A paired-end sequencing strategy extends reliable read length by separately sequencing each end of a molecule (sometimes referred to as a "tag" end) that comprise a fragment of an original template nucleic acid molecule at each end joined in the center by a linker sequence. The original positional relationship of the template fragments is known and thus the data from the sequence reads may be re-combined into a single read having a longer

IDT Exhibit 1019.0194

high quality read length. Further examples of paired-end sequencing embodiments are described in US Patent Application Serial No. 11/448,462, titled "Paired end sequencing", filed June 6, 2006, and in US Provisional Patent Application Serial No. 60/026,319, titled "Paired end sequencing", filed February 5, 2008, each of which is hereby incorporated by reference herein in its entirety for all purposes.

Some examples of SBS apparatus may implement some or all of the methods described above may include one or more of a detection device such as a charge coupled device (i.e. CCD camera), a microfluidics chamber or flow cell, a reaction substrate, and/or a pump and flow valves. Taking the example of pyrophosphate based sequencing, embodiments of an apparatus may employ a chemiluminescent detection strategy that produces an inherently low level of background noise.

In some embodiments, the reaction substrate for sequencing may include what is referred to as a PicoTiterPlate® array (also referred to as a PTP® plate) formed from a fiber optics faceplate that is acid-etched to yield hundreds of thousands of very small wells each enabled to hold a population of substantially identical template molecules. In some embodiments, each population of substantially identical template molecule may be disposed upon a solid substrate such as a bead, each of which may be disposed in one of said wells. For example, an apparatus may include a reagent delivery element for providing fluid reagents to the PTP plate holders, as well as a CCD type detection device enabled to collect photons of light emitted from each well on the PTP plate. Further examples of apparatus and methods for performing SBS type sequencing and pyrophosphate sequencing are described in US Patent No 7,323,305 and US Patent Application Serial No. 11/195,254 both of which are incorporated by reference above.

In addition, systems and methods may be employed that automate one or more sample preparation processes, such as the emPCR™ process described above. For example, microfluidic technologies may be employed to provide a low cost, disposable solution for generating an emulsion for emPCR processing, performing PCR Thermocycling operations, and enriching for successfully prepared populations of nucleic acid molecules for sequencing. Examples of microfluidic systems for sample preparation are described in U.S. Provisional Patent Application Serial No. 60/915,968, titled "System and Method for Microfluidic Control of Nucleic Acid amplification and Segregation", filed May 4, 2007, which is hereby incorporated by reference herein in its entirety for all purposes.

Also, the systems and methods of the presently described embodiments of the invention may include implementation of some design, analysis, or other operation using a

IDT Exhibit 1019.0195

computer readable medium stored for execution on a computer system. For example, several embodiments are described in detail below to process detected signals and/or analyze data generated using SBS systems and methods where the processing and analysis embodiments are implementable on computer systems.

An exemplary embodiment of a computer system for use with the presently described invention may include any type of computer platform such as a workstation, a personal computer, a server, or any other present or future computer. Computers typically include known components such as a processor, an operating system, system memory, memory storage devices, input-output controllers, input-output devices, and display devices. It will be understood by those of ordinary skill in the relevant art that there are many possible configurations and components of a computer and may also include cache memory, a data backup unit, and many other devices.

Display devices may include display devices that provide visual information, this information typically may be logically and/or physically organized as an array of pixels. An interface controller may also be included that may comprise any of a variety of known or future software programs for providing input and output interfaces. For example, interfaces may include what are generally referred to as "Graphical User Interfaces" (often referred to as GUI's) that provide one or more graphical representations to a user. Interfaces are typically enabled to accept user inputs using means of selection or input known to those of ordinary skill in the related art.

In the same or alternative embodiments, applications on a computer may employ an interface that includes what are referred to as "command line interfaces" (often referred to as CLI's). CLI's typically provide a text based interaction between an application and a user. Typically, command line interfaces present output and receive input as lines of text through display devices. For example, some implementations may include what are referred to as a "shell" such as Unix Shells known to those of ordinary skill in the related art, or Microsoft Windows Powershell that employs object-oriented type programming architectures such as the Microsoft .NET framework.

Those of ordinary skill in the related art will appreciate that interfaces may include one or more GUI's, CLI's or a combination thereof.

A processor may include a commercially available processor such as a Centrino®, Core™ 2, Itanium® or Pentium® processor made by Intel Corporation, a SPARC® processor made by Sun Microsystems, an Athalon™ or Opteron™ processor made by AMD corporation, or it may be one of other processors that are or will become available. Some

embodiments of a processor may include what is referred to as Multi-core processor and/or be enabled to employ parallel processing technology in a single or multi-core configuration. For example, a multi-core architecture typically comprises two or more processor "execution cores". In the present example each execution core may perform as an independent processor that enables parallel execution of multiple threads. In addition, those of ordinary skill in the related will appreciate that a processor may be configured in what is generally referred to as 32 or 64 bit architectures, or other architectural configurations now known or that may be developed in the future.

A processor typically executes an operating system, which may be, for example, a Windows®-type operating system (such as Windows® XP or Windows Vista®) from the Microsoft Corporation; the Mac OS X operating system from Apple Computer Corp. (such as 7.5 Mac OS X v10.4 "Tiger" or 7.6 Mac OS X v10.5 "Leopard" operating systems); a Unix® or Linux-type operating system available from many vendors or what is referred to as an open source; another or a future operating system; or some combination thereof. An operating system interfaces with firmware and hardware in a well-known manner, and facilitates the processor in coordinating and executing the functions of various computer programs that may be written in a variety of programming languages. An operating system, typically in cooperation with a processor, coordinates and executes functions of the other components of a computer. An operating system also provides scheduling, input-output control, file and data management, memory management, and communication control and related services, all in accordance with known techniques.

System memory may include any of a variety of known or future memory storage devices. Examples include any commonly available random access memory (RAM), magnetic medium such as a resident hard disk or tape, an optical medium such as a read and write compact disc, or other memory storage device. Memory storage devices may include any of a variety of known or future devices, including a compact disk drive, a tape drive, a removable hard disk drive, USB or flash drive, or a diskette drive. Such types of memory storage devices typically read from, and/or write to, a program storage medium (not shown) such as, respectively, a compact disk, magnetic tape, removable hard disk, USB or flash drive, or floppy diskette. Any of these program storage media, or others now in use or that may later be developed, may be considered a computer program product. As will be appreciated, these program storage media typically store a computer software program and/or data. Computer software programs, also called computer control logic, typically are stored in

system memory and/or the program storage device used in conjunction with memory storage device.

In some embodiments, a computer program product is described comprising a computer usable medium having control logic (computer software program, including program code) stored therein. The control logic, when executed by a processor, causes the processor to perform functions described herein. In other embodiments, some functions are implemented primarily in hardware using, for example, a hardware state machine. Implementation of the hardware state machine so as to perform the functions described herein will be apparent to those skilled in the relevant arts.

Input-output controllers could include any of a variety of known devices for accepting and processing information from a user, whether a human or a machine, whether local or remote. Such devices include, for example, modem cards, wireless cards, network interface cards, sound cards, or other types of controllers for any of a variety of known input devices. Output controllers could include controllers for any of a variety of known display devices for presenting information to a user, whether a human or a machine, whether local or remote. In the presently described embodiment, the functional elements of a computer communicate with each other via a system bus. Some embodiments of a computer may communicate with some functional elements using network or other types of remote communications.

As will be evident to those skilled in the relevant art, an instrument control and/or a data processing application, if implemented in software, may be loaded into and executed from system memory and/or a memory storage device. All or portions of the instrument control and/or data processing applications may also reside in a read-only memory or similar device of the memory storage device, such devices not requiring that the instrument control and/or data processing applications first be loaded through input-output controllers. It will be understood by those skilled in the relevant art that the instrument control and/or data processing applications, or portions of it, may be loaded by a processor in a known manner into system memory, or cache memory, or both, as advantageous for execution.

Also a computer may include one or more library files, experiment data files, and an internet client stored in system memory. For example, experiment data could include data related to one or more experiments or assays such as detected signal values, or other values associated with one or more SBS experiments or processes. Additionally, an internet client may include an application enabled to accesses a remote service on another computer using a network and may for instance comprise what are generally referred to as "Web Browsers". In the present example some commonly employed web browsers include Microsoft® Internet

IDT Exhibit 1019.0198

Explorer 7 available from Microsoft Corporation, Mozilla Firefox® 2 from the Mozilla Corporation, Safari 1.2 from Apple Computer Corp., or other type of web browser currently known in the art or to be developed in the future. Also, in the same or other embodiments an internet client may include, or could be an element of, specialized software applications enabled to access remote information via a network such as a data processing application for SBS applications.

A network may include one or more of the many various types of networks well known to those of ordinary skill in the art. For example, a network may include a local or wide area network that employs what is commonly referred to as a TCP/IP protocol suite to communicate. A network may include a network comprising a worldwide system of interconnected computer networks that is commonly referred to as the internet, or could also include various intranet architectures. Those of ordinary skill in the related arts will also appreciate that some users in networked environments may prefer to employ what are generally referred to as "firewalls" (also sometimes referred to as Packet Filters, or Border Protection Devices) to control information traffic to and from hardware and/or software systems. For example, firewalls may comprise hardware or software elements or some combination thereof and are typically designed to enforce security policies put in place by users, such as for instance network administrators, etc.

b. Embodiments of the presently described invention

As described above, the presently described invention comprises associating one or more embodiments of a UID element having a known and identifiable sequence composition with a sample, and coupling the embodiments of UID element with template nucleic acid molecules from the associated samples. The UID coupled template nucleic acid molecules from a number of different samples are pooled into a single "Multiplexed" sample or composition that can then be efficiently processed to produce sequence data for each UID coupled template nucleic acid molecule. The sequence data for each template nucleic acid is de-convoluted to identify the sequence composition of coupled UID elements and association with sample of origin identified. For example, a multiplexed composition may include representatives from about 384 samples, about 96 samples, about 50 samples, about 20 samples, about 16 samples, about 10 samples, or other number of samples. Each sample may be associated with a different experimental condition, treatment, species, or individual in a research context. Similarly, each sample may be associated with a different tissue, cell, individual, condition, or treatment in a diagnostic context. Those of ordinary skill in the

IDT Exhibit 1019.0199

related art will appreciate that the numbers of samples listed above are for the purposes of example and thus should not be considered limiting.

Typically, systems and methods are employed for processing samples to generate sequence data as well as for interpretation of the sequence data. Figure 1 provides an illustrative example of sequencing instrument 100 employed to execute sequencing processes using reaction substrate 105 that for instance may include the PTP® plate substrate described above. Also illustrated in Figure 1 is computer 130 that may for instance execute system software or firmware for processing as well as perform analysis functions. In the example of Figure 1, computer 130 may also store application 135 in system memory for execution, where application 135 may perform some or all of the data processing functions described herein. It will also be understood that application 135 may be stored on other computer or server type structures for execution and perform some or all of its functions remotely communicating over networks or transferring information via standard media. For instance, processed target molecules in a multiplex sample may be loaded onto reaction substrate 105 by user 101 or some automated embodiment then sequenced in a massively parallel manner using sequencing instrument 100 to produce sequence data representing the sequence composition of each target molecule. Importantly, user 101 may include any user such as independent researcher, university, or corporate entity. In the present example, sequencing instrument 100, reaction substrate 105, and/or computer 130 may include some or all of the components and characteristics of the embodiments generally described above.

In preferred embodiments, the sequence composition of each UID element is easily identifiable and resistant to introduced error from sequencing processes. Some embodiments of UID element comprise a unique sequence composition of nucleic acid species that has minimal sequence similarity to a naturally occurring sequence. Alternatively, embodiments of a UID element may include some degree of sequence similarity to naturally occurring sequence.

Also, in preferred embodiments the position of each UID element is known relative to some feature of the template nucleic acid molecule and/or adaptor elements coupled to the template molecule. Having a known position of each UID is useful for finding the UID element in sequence data and interpretation of the UID sequence composition for possible errors and subsequent association with the sample of origin.

For example, some features useful as anchors for positional relationship to UID elements may include, but are not limited to the length of the template molecule (i.e. the UID element is known to be so many sequence positions from the 5' or 3' end), recognizable sequence

IDT Exhibit 1019.0200

markers such as a Key element (described in greater detail below) and/or one or more primer elements positioned adjacent to a UID element. In the present example, The Key and primer elements generally comprise a known sequence composition that typically does not vary from sample to sample in the multiplex composition and may be employed as positional references for searching for the UID element. An analysis algorithm implemented by application 135 may be executed on computer 130 to analyze generated sequence data for each UID coupled template to identify the more easily recognizable Key and/or primer elements, and extrapolate from those positions to identify a sequence region presumed to include the sequence of the UID element. Application 135 may then process the sequence composition of the presumed region and possibly some distance away in the flanking regions to positively identify the UID element and its sequence composition.

Also, as will be described in greater detail below in some embodiments the sequence data generated from each Key and/or one or more primer elements may be analyzed to determine a measure of the relative error rate for the sequencing run. The measure of error rate may then be employed in the analysis of the sequence data generated for the UID element. For example, if the error rate is excessive and is above a predetermined threshold it may also be assumed that a similar rate of error exists in the sequence data generated for the UID element, and thus the sequence data for the entire template may be filtered out as suspect. Further, in embodiments where a UID element is coupled to each end of a linear template molecule an error rate may be established for each end and asymmetrically analyzed. Importantly, it will be appreciated that in some embodiments, particularly sequencing technology capable of producing "long" read lengths (i.e. of about 100 base pairs or greater) the error rate in the sequence data may differ between the 5' end and the 3' end.

In preferred embodiments, a UID element is associated with an adaptor enabled to operatively couple with the end of a template nucleic acid molecule. In typical high throughput sequencing applications it is desirable that the template nucleic acid molecules are linear where an adaptor may be coupled to each end. Figures 2A and 2B provide illustrative examples of embodiments of adaptor composition for various applications comprising one or more UID elements. It will, however, be appreciated that various adaptor configurations may be employed for different amplification and sequencing strategies. Figure 2A provides an illustrative example of adaptor element 200 that comprises an embodiment of an adaptor amenable for use with amplification and sequencing of Genomic Libraries. It will also be appreciated that adaptor element 200 may also be amenable for libraries of template molecules independently amplified with target specific sequences independently of the

IDT Exhibit 1019.0201

adaptor element described herein. Adaptor element 200 comprises several components that include primer 205, key 207, and UID 210. Also, Figure 2B provides an illustrative example of one embodiment of adaptor 220 amenable for use with amplification and sequencing of Amplicons. Adaptor element 220 comprises several similar components to adaptor 200 that include primer 205, key 207, UID 210, with the addition of target specific element 225. It will be appreciated that the relative arrangement of components provided in Figures 2A and 2B are for illustrative purposes and should not be considered limiting.

In some alternative embodiments, the UID 210 elements are not associated with adaptor elements as described above. Rather, the UID 210 elements may be considered separate elements that may be independently coupled to an already adapted template molecule, or non-adapted template molecule. This strategy may be useful in some circumstances to avoid negative effects associated with a particular step or assay. For example, it may be advantageous in some embodiments to ligate the UID 210 elements to each population of substantially identical template molecules after copies have been produced from an amplification step. By coupling the UID elements to the adapted template molecules post-amplification, errors introduced by the amplification method are avoided. In the present example, PCR amplification methods that employ polymerases are known to have a certain rates of introduced error based, at least in part, upon the type of polymerase or polymerase blends (i.e. a blend may include a mixture of what may be referred to as a "high fidelity" polymerase and a polymerase with "proof reading" capability) employed and the number of cycles of amplification.

It will also be appreciated that multiple embodiments of adaptor 200 or 220 may be employed with each template molecule, such as one embodiment of adaptor 200 or 220 at each end of a linear template molecule prepared for sequencing. However, in some embodiments the positional arrangement of elements within adaptor 200 or 220 may be reversed (i.e. the elements of adaptor 200 or 220 are in a palindromic arrangement from the example illustrated in Figure 2A or 2B) at the 3' end relative the arrangement of elements in adaptor 200 or 220 at the 5' end. For example, an embodiment of element 220 may be positioned on each end of substantially every template molecule from a library of amplicons in a multiplex composition, thus 2 embodiments of UID 210 may be employed in a combinatorial manner for identification which will be discussed in greater detail below.

Primer 205 may include a primer species (or a primer of a primer pair) such as is described above with respect to emulsion PCR embodiments (i.e. Primer A and Primer B). Also, primer 205 may include a primer species employed for an SBS sequencing reaction

IDT Exhibit 1019.0202

also as described above. Further, primer 205 may include what is referred to as a bipartite PCR/sequencing primer useable for both the emulsion PCR and SBS sequencing processes. Key 207 may include what may be referred to as a "discriminating key sequence" that refers to a short sequence of nucleotide species such as a combination of the four nucleotide species (i.e., A, C, G, T). Typically, key 207 may employed for quality control of sequence data, where for example key 207 may be located immediately adjacent primer 205 or within close proximity and include one of each of the four nucleotide species in a known sequence arrangement (i.e. TCAG). Therefore, the fidelity of the sequencing method should be represented in the sequence data for each of the 4 nucleotide species in key 207 and may pass quality control metrics if each of the 4 nucleotide species is faithfully represented. For example, an error for one of the nucleotide species represented in the sequence data generated from key 207 could indicate a problem in the sequencing process associated with that nucleotide species. Such error may be from mechanical failure of one or more components of sequencing instrument 100, low quality or supply of reagent, operating script error, or other source of systematic type error that may occur. Thus, if such systematic type error is detected in key 207 that sequence data generated for the run of that template molecule may not pass quality metrics and will typically be rejected.

The same discriminating sequence for key 207 can be used for an entire library of DNA fragments, or alternatively different sequence compositions may be associated with portions of the library for different purposes. Further examples of primer and key elements associated with primer 205 and key 207 are described in U.S. Patent Application Serial No. 10/767,894, incorporated by reference above.

Target specific element 225 includes a sequence composition that specifically recognizes a region of a genome. For example, Target specific element 225 may be employed as a primer sequence to amplify and produce amplicon libraries of specific targeted regions for sequencing such as those found within genomes, tissue samples, heterogeneous cell populations or environmental samples. These can include, for example, PCR products, candidate genes, mutational hot spots, evolutionary or medically important variable regions. It could also be used for applications such as whole genome amplification with subsequent whole genome sequencing by using variable or degenerate amplification primers. Further examples describing the use of target specific sequences with bipartite primers are described in U.S. Patent Application Serial No. 11/104,781, titled "Methods for determining sequence variants using ultra-deep sequencing", filed April 12, 2005, which is hereby incorporated by reference herein in its entirety for all purposes.

IDT Exhibit 1019.0203

Some embodiments of UID 210 may be particularly amenable for use with relatively small numbers of sample associations in a multiplex sample. In particular, when there are only a small number of associations to identify in a multiplex sample, each sample is associated with a distinct implementation of UID 210 comprising a sequence composition that is sufficiently unique from each other as to enable easy detection and correction of introduced error. In some embodiments, groups of compatible UID 210 sequence elements are clustered into "sets" as will be described in greater detail below. For example, a set of UID 210 elements may include 14 members that may be employed to uniquely identify up to 14 associations with samples, where each member is associated with a single sample.

It will be appreciated that as the number of associations to identify grows, it becomes increasingly difficult to design distinct embodiments of UID 210 for each association that meet the design criteria and desired characteristics. In such cases, it may be advantageous to employ multiple UID 210 elements combinatorially to uniquely associate the template molecules with their sample of origin, where one embodiment of UID 210 may be positioned at each end of a linear template molecule. For example, the number of associations to identify between the sequence data generated from template molecules and the sample of origin may become too large to accommodate given the necessary design parameters and characteristics of UID 210. In particular, it is undesirable in many embodiments to employ a distinct UID element for each association when the number a samples would require a sequence length for UID 210 that is undesirably long for the design criteria that includes a specific number of flow cycle iterations and number of sequence positions taken up by the UID element. In the present example, in embodiments of sequencing technology that generate "long" read lengths UID 210 may comprise up to 10 sequence positions. Alternatively, other embodiments of sequencing technology may generate relatively short read lengths of about 25-50 sequence positions, and thus it is desirable that UID 210 is short in order to optimize the read length for the template molecule. In the present example, UID 210 may be designed for short read lengths comprising up to 4 sequence positions, up to 6 sequence positions, or up to 8 sequence positions, depending, at least in part, upon the application.

As described above, embodiments for design and implementation of UID 210 amenable for both small and large numbers of associations is to employ a "set" of UID 210 elements each meeting the preferred design criteria and characteristics. In some applications, such as the design of UID 210 elements with sequence composition that enable accurate error detection and correction features it is desirable to use the "set" strategy presently described.

IDT Exhibit 1019.0204

For example, as will be described in greater detail below the sequence composition for the UID elements in a set must be sufficiently distinct from each other in order to enable error detection and correction thereby limiting the compatible members available for a particular set. However, UID 210 members from multiple sets may be combinatorially employed with a template molecule where the members of each set are located at different relative positions and are thus easily interpretable.

In order to overcome the problems of a large number of associations to identify described above, two or more members from a set of UID 210 elements may be employed in a combinatorial manner. For example, a set of UID 210 elements may include 10, 12, 14, or other number of members comprising a 10-mer sequence length. In some embodiments, two UID 210 elements may be associated with each template molecule and used combinatorially to identify up to 144 different associations (i.e. 12 UID members for use with element 1 multiplied by 12 UID members for use with element 2 results in 144 possible combinations of UID elements 1 and 2 that may be employed to uniquely identify an association).

Those of ordinary skill in the related art will appreciate that alternative embodiments may be employed where each UID 210 element associated with a template molecule may include a subset of the total number of UID members from the set (i.e. use a portion of the members of the set). In other words, of the 12 members of a complete set, only 8 may be employed at one element position. There are a number of reasons why it may be desirable to use a subset of UID members that includes having a need for a smaller number of associations to identify (i.e. smaller number of combinations), physical or practical experimental conditions such as equipment or software limitations, or preferred combinations of UID members of a set in element positions. For instance, a first element may employ all 12 UID members from a set and a second element may employ a subset of 8 UID members from the same or different set yielding 96 possible combinations.

UID 210 elements used in combinatorial strategies may be configured in a variety of positional arrangements relative to the position of the template molecule. For example, a strategy that utilizes 2 UID 210 elements combinatorially to identify the association of each template molecule with its sample of origin may include a UID element positioned at each end of a linear template molecule (i.e. one UID 210 element at the 5' end and another at the 3' end). In the present example, each UID 210 element may be associated with an adaptor element, such as adaptor 200 or 220, employed in a target specific amplicon or genomic library sequencing strategy as discussed above. Thus, the sequence data associated with a template molecule would include the sequence composition of a UID element at each end of

IDT Exhibit 1019.0205

the amplicon. The combination of the UID elements may then be used to associate the sequence data with the sample of origin of the template molecule.

In some alternative embodiments, a UID 210 element may be incorporated in an adaptor element at each end of a linear template molecule as described above. However, the read length of the template molecule may be greater than the ability of the sequencing technology to handle. In such a case, the template molecule may be sequenced from each end independently (i.e. a separate sequencing run for each end), where the UID 210 element associated with the end may be employed as a single UID 210 identifier.

In addition it may be desirable in some embodiments to assign more that one UID 210 element per sample, or more than one combinations of UID 210 elements. Such a strategy may provide redundancy to protect against possible unintended biases introduced by various source, which could include the UID 210 element itself. For example, a sample with a population of template molecules may be sub-divided in sub-samples each using a distinctive UID 210 element for the association. In such a case, the redundancy of the different UID 210 elements for the same population of template molecules from a sample provides for greater confidence that the correct associations will be identified or if the error is too great to make a correct identification of the association with confidence.

As generally described above, embodiments of the presently described invention include one or more UID 210 elements operatively coupled to each template molecule for the purpose of identifying the association between the template molecule and the sequence data generated therefrom with a sample of origin. One or more embodiments of a UID element may be operatively coupled to one or more components of an adaptor and a template molecule using a variety of methods known in the art that include but are not limited to ligation techniques. Methods for ligating nucleic acid molecules to one another are generally known in the art and include employing a ligase enzyme for what is referred to as sticky end or blunt end ligation. Further examples of coupling adaptor elements to template molecules using ligation as described in U.S. Patent Application Serial No. 10/767,894, titled "Method for preparing single-stranded DNA libraries", filed January 28, 2004; and U.S. Provisional Patent Application Serial No. 60/031,779, titled "System and Method for Improved Processing of Nucleic Acids for Production of Sequencable Libraries" filed February 27, 2008, each of which is hereby incorporated by reference herein in its entirety for all purposes). For example, a large template nucleic acid or whole genomic DNA sample may be fragmented by mechanical (i.e. nebulization, sonication) or enzymatic means (i.e. DNase I), the resulting ends of each fragment may be polished for compatibility with adaptor

IDT Exhibit 1019.0206

elements (i.e. polishing using what is referred to as an exonuclease, such as BAL32 nuclease or Mung Bean nuclease), and each fragment may be ligated to one or more adaptor elements (i.e. using T4 DNA ligase). In the present example, each adaptor element is directionally ligated to the fragment such as for instance by selective binding between the 3' end of the adaptor and the 5' end of the fragment.

In some embodiments, UID 210 elements may be provided to user 101 in the form of a kit, where the kit could include adaptors comprising incorporated UID 210 elements as illustrated in Figures 2A and 2B. Or, the kit could include UID 210 as independent elements that enable user 101 to incorporate as they desire.

As described above, embodiments of UID 210 should comprise a number of preferred characteristics or design criteria that include but are not limited to a) each UID element comprises a minimal sequence length requiring a minimal number of synthesis or flow cycles, b) each UID element comprises sequence distinctiveness, c) each UID element comprises resistance to introduced error, and d) each UID element does not interfere with amplification methods (such as PCR, or cloning into vectors).

Also, some embodiments of UID element design may also consider physical characteristics or design criteria of nucleic acids that include some or all of i) UID sequence composition selected to resist formation of what are referred to as "hairpins" (also referred to as a "hairpin loop" or "stem loop") and "primer dimers"; ii) UID elements comprise preferred melting temperature (i.e. 40°C) and/or Gibbs free energy (i.e. ΔG cutoff of -1.5) characteristics. Aspects of some of the desirable characteristics and their impact on UID design are described in greater detail below.

One important characteristic of a UID element is that it should include a minimal number of bases or sequence positions required to satisfy the needs of other characteristic requirements. For example, each UID element should comprise the minimum sequence length required to uniquely identify a desired number of associations between the template molecule/sequence data and their samples of origin. A desired number of associations may include identification of template molecules/sequence data associated with at least 12 different samples, at least 96 different samples, at least 384 different samples, or a greater number of samples that may be contemplated in the future. In other words the sequence length of the UID should be no longer than necessary in order to conserve the number of positions (i.e. what may be referred to as "sequence real estate") of the read length for the template molecule. Further, the minimum sequence length should consume or require a minimum number of flow cycles of the set of nucleotide species to generate the sequence

IDT Exhibit 1019.0207

data for each UID element. Minimizing the number of nucleotide species flow cycles required to generate sequence data for the UID elements provides advantages in reagent cost, instrument usage (i.e. processing time), data quality, and read length. For instance, each additional flow cycle increases the probability of introducing CAFIE error, and reagent usage. In the present example, it is preferable that each 10-mer UID element require only 5 nucleotide species flow cycles to generate sequence data for each UID element.

Another important characteristic includes sequence distinctiveness of each UID element. The term "sequence distinctiveness" as used herein generally refers to a distinguishable difference between a plurality of UID sequences such that each sequence is easily recognizable from every other UID sequence that is the subject of comparison. In particular each UID element needs to comprise a measure of sequence distinctiveness that enables easy detection of introduced error and correction of some or all of the error. Further, it is generally preferable that each UID element be free of repetitive sequence composition and should not include a sequence composition recognized by restriction enzymes. In other words it is undesirable for UID elements to include consecutive monomers having the same composition of nucleotide species. For example, preferred embodiments of the sequence distinctiveness of each UID element enable detection of up to 3 sequence positions with introduced errors and correction of up to 2 sequence positions with introduced errors in a 10-mer element (i.e. 10 total sequence positions). Those of ordinary skill will appreciate that the introduced error may include what are referred to as "insertions", "deletions", "substitutions", or some combination thereof (i.e. a combination of an insertion and deletion at the same sequence position will appear to be a substitution and would be counted as a single error event). Also, the level of error detection and correction may depend, at least in part, upon the sequence length of the UID element. Further, introduced errors outside (i.e. upstream or downstream) of UID 210 may have effects on the interpretation of sequence composition for UID 210. This will be discussed further below in the context of decoding or analysis of sequence data for UID identification.

A further characteristic that is also desirable comprises resistance to introduced error. For example, monomer repeats in nucleic acid sequence such as that of the template molecule or other sequence elements may cause errors in a sequence read. The error may include an over or under representation or call of the number of repeated monomers. It is therefore desirable that the UID elements do not begin or end with the same nucleotide species as the adjacent monomer of a neighboring sequence element (i.e. creating monomer repeats between sequence elements or components). In the present example, a neighboring sequence

IDT Exhibit 1019.0208

element, such as key 207 illustrated in Figures 2A and 2B, may end with a "G" nucleotide species. Therefore, a UID element such as UID 210, should not begin with the same "G" nucleotide species to avoid the increased possibility introduced error from the repeated "G" species.

Another source of error that is particularly relevant in SBS contexts, include what are referred to as "carry forward" or "incomplete extension" effects (sometimes referred to as CAFIE effects). For example, a small fraction of template nucleic acid molecules in each amplified population of a nucleic acid molecule from a sample (i.e. a population of substantially identical copies amplified from a nucleic acid molecule template) loses or falls out of phasic synchronism with the rest of the template nucleic acid molecules in the population (that is, the reactions associated with the fraction of template molecules either get ahead of, or fall behind, the other template molecules in the sequencing reaction run on the population) . Additional description of CAFIE mechanisms and methods of correcting CAFIE error are further described in PCT Application Serial No US2007/004187, titled "System and Method For Correcting Primer Extension Errors in Nucleic Acid Sequence Data", filed February 15, 2007, which is hereby incorporated by reference herein in its entirety for all purposes.

Also, it will be appreciated that some types of error may occur at higher frequency than other types and/or have greater consequences than other types of error. For example, deletion error may have more significant impact than substitution error. It is therefore advantageous to design each UID element so that it is weighted more heavily to deal with the more frequent or more deleterious types of error.

As stated previously, it is not typically desirable to randomly or non-selectively design the sequence composition of UID elements. An illustrative example of two improperly designed UID elements and the potential for problems with error detection/correction using such UID elements is presented in Table 1.

Table 1:

UID Element 1	Generated UID Sequence	UID Element 2
A*C*TGA (SEQ ID NO: 1)	A*G*T*G*A (SEQ ID NO: 2)	AG*C*GA (SEQ ID NO: 3)

In the example of table 1, it is apparent that the UID sequence represented as generated UID sequence contains an error (i.e. the presence of at least one error is detected) if either UID element 1 or 2 is the original sequence element. However, it is not clear from the

IDT Exhibit 1019.0209

sequence composition of the Generated UID sequence whether UID element 1 or UID element 2 was the actual UID element because a single error in either could result in the generated sequence. In other words, it is possible that one error was introduced in UID element 1 transforming the "C" nucleotide species at the second position to a "G" species. It is also possible that one error was introduced in UID element 2 transforming the "C" nucleotide species at the third position to a "T" species. Given the sequence information, the error is detected but it is not possible to infer which UID element was the original element and thus cannot be corrected. Therefore, the association of the generated UID sequence with either UID element 1 or 2 cannot be positively made, and thus the sample of origin for the template molecule coupled to one of the UID elements cannot be identified and the generated sequence information may need to be thrown out. In other words, the design of UID elements 1 and 2 are not sufficiently distinct from each other to recover from the described type of introduced error.

The potential result of poor UID design is further exemplified in Table 2.

Table 2:

UID Element 1	UID Element 2
CTACC (SEQ ID NO: 4)	CTGCC (SEQ ID NO: 5)

The example of Table 2 provides an even clearer picture of the potential consequences where a substitution event in UID element 1 of an A nucleotide species at the third position to a G nucleotide species, which is one of the most common types of error introduced by PCR processes, results in an exact match with the sequence composition of UID 210 element. Thus the poor UID 210 design results in an undetectable error that would likely result in the mis-assignment of the sequence data to a sample of origin.

Various methods may be employed to design UID elements comprising sequence composition that meets the necessary design criteria. Also, application 135 illustrated in Figure 1 may be employed for designing UID 210 using some or all of the methods described herein. For example, "Brute Force" methods may be employed that compute every possible sequence composition for a given length and the possible conflicts with other sequence composition given a set of parameters associated with the design criteria. In the present example, the sequence composition of 10 mer UID elements may be computed for detection of up to 3 sequence positions with introduced errors and correction of up to 2 sequence positions with introduced errors.

IDT Exhibit 1019.0210

Design of a preferred sequence composition for members of a set of UID 210 elements meeting the most stringent design criteria given the characteristics described above presents a computational challenge. Mathematical methods known to those of skill in the art may be applied to compute the possible sequence composition for members of a set given the design constraints. For example, mathematical transformations of all possible combinations of sequence composition may be computed given the design constraints to generate what may be referred to as "Error Balls" or "Error Clouds" to determine the potential compatibility of each UID element with the other members in a set. Compatibility of sequence composition for potential UID elements may be visually illustrated as non-overlapping error balls. For example, Figure 3 provides an illustrative representation of what may be referred to as "space potential" for computed error balls for UID 310, UID 320, UID 330, UID 340, and UID 350 comprising some or all of the design criteria described above such as number of flow cycles, and sequence length requirements. As illustrated in Figure 3 the error balls for UID 310, UID 320, and UID 330 do not overlap and thus represent sequence composition of compatible UID 210 elements. Further, UID 340 overlaps with UID 320 and UID 350 representing a sequence composition for a UID element that is not compatible. However UID 340 does not overlap with UID 310 and UID 330 and thus represents compatible sequence composition for each non-overlapping UID element.

Alternatively, a more computationally efficient approach may be employed that uses what is referred to in the art as "Dynamic Programming" techniques. The term "Dynamic Programming" as used herein generally refers to methods for solving problems that comprise overlapping sub-problems and optimal structure. Dynamic programming techniques are typically substantially more computationally efficient than methods with no *a priori* knowledge.

Some embodiments of dynamic programming technique include computing what may be referred to as the "minimum edit distance" for strings of characters such as strings of nucleic acid species. In other words, each UID member element in a set may be considered a string of characters representing the nucleic acid species composition. The term "minimum edit distance" as used herein generally refers to the minimum number of point mutations required to change a first string into a second string. Further, the term "point mutation" as used herein generally refers to and includes a change of character composition at a location in a string referred to as a substitution of a character for another in a string; an insertion of a character into a string; or a deletion of a character from a string. For example, the minimum edit distance may be computed for each potential member of a set of UID 210 elements

IDT Exhibit 1019.0211

against all other members of the set. Subsequently the minimum edit distances may be compared and members of the set of UID 210 elements selected based, at least in part, upon each member of the set having a sufficiently high minimum edit distance from all other members to meet the specified criteria. Systems and methods for computing minimum edit distance are well known to those of ordinary skill in the related art and may be implemented in a number of ways.

Another important aspect of the presently described invention is directed to the analysis of sequence data to "decode" or identify the UID 210 sequence elements within the data. In some embodiments an algorithm may be implemented in computer code as application 135 that processes the sequence data from each run and identify UID 210 as well as perform any error detection or corrections functions. It is important to recognize that methods of error detection and correction in strings of information have been employed in the computer arts particularly in the area of electronically stored and transmitted data. For example, the problem of "inversion" of bits of data from one form into another occurs when data is transmitted over networks or stored in electronic media. The inversion of bits presents a problem with respect to the integrity of stored or transmitted data and is analogous to the presently described substitution type of error. Methods of detection and correction of inversion error is described in J. F. Wakerly, "Detection of unidirectional multiple errors using low cost arithmetic codes," IEEE Trans. Comput., vol. C-24, pp. 210-212, Feb. 1975.; and J. F. Wakerly, Error Detecting Codes, Self-Checking Circuits and Applications. Amsterdam, The Netherlands: North-Holland, 1978, both of which are hereby incorporated by reference herein in their entireties for all purposes.

However, the methods of detecting and correcting inversion error described above are not applicable to the problem of error detection and correction in sequence data and more specifically errors in UID elements. Importantly, the problem in sequence data is substantially more complex because it deals with the problems of substitutions and deletions as well as substitutions that create phasing problems and complicate the interpretation of information at each sequence position.

As described above, UID 210 may be located at a known position relative to other easily identifiable elements such as primer 205, key 207, the 5' or 3' end of the sequence, etc. However, just as introduced error within UID 210 has deleterious effects, error outside of the region of the UID 210 element may also affect the efficiency of identifying each UID 210 element. Further, some types of error outside of the region defined by UID 210 may contribute to and count as errors within UID 210 sequence. For example, insertion events

may occur and be represented in the sequence data preceding (i.e. upstream of) UID 210 element that may be difficult to interpret. In the present example, an insertion event could include the insertion of one or more G nucleotide species bases at the end of key 207 comprising a TCAG sequence composition as may occur when a nucleotide species at a sequence position is "overcalled". However, an application that interprets the data will not know that it is an insertion event and cannot rule out the possibility of a substitution event that provided a G nucleotide in place of a different nucleotide species at the first sequence position of UID 210. In other words, the error outside of UID 210 will force the algorithm to decide if the error is an insertion that shifts where it should look for the first sequence position of UID 210 or whether it is a substitution event.

Continuing the example from above, an algorithm or user may look for the UID 210 element immediately adjacent to another known element such as key 207 as illustrated in Figures 2A and 2B, but the insertion of one base between key 207 and UID 210 may typically be assigned as belonging to UID 210 (counts as a first insertion error). Additionally, the algorithm or user expects UID 210 to be a certain length (i.e. 10 sequence positions) and thus truncates the last sequence position of the actual UID element because of the first insertion (counts as a second deletion error). Thus, it is clear that errors outside of the UID region can have substantial effect on finding and interpreting the sequence composition of UID 210.

In some embodiments, errors outside of the region defined by UID 210 may be particularly troublesome at the 3' end of a nascent molecule. For example, some embodiments of SBS sequence from 5' to 3' ends (i.e. adding nucleotide species to 3' end of nascent molecule) where cumulative errors (such as CAFIE type error described above) and the rate of introduced error may be increasingly higher as the sequence run gets longer at the 3' end. Thus, it may be more practical and effective to use certain assumptions rather than stringent criteria to identify UID 210. Also as described above, assumptions used for the 5' may be different than assumptions employed for the 3' end and may be referred to as "Asymmetric". For example, it may be assumed that there will never be more than 3 sequence position errors present at the 5' end which would be consistent with empirical evidence. However, in the present example at the 3' end it may be assumed that there will never be more than 4 sequence position errors due to the increased possibility of error at the 3' end. Because of the asymmetric difference in detectable error at each end, it may also be inferred that the amount of that error that is correctable may also be different. In the present example, the correctable error at the 5' end may be 2 sequence positions as described above, however the correctable error at the 3' end may only be 1 sequence position. Also, further

IDT Exhibit 1019.0213

assumptions may be employed at the 3' end that may not be employed for the 5' end. Such an assumption could include the existence of one or more "no called" positions in close proximity to UID 210.

In the present example, an embodiment of adaptor element 200 or 220 is present at the 3' end of a template nucleic acid in a palindromic arrangement to that illustrated in Figure 2A or 2B (as described above). It will be appreciated however, that the present example refers to a difference in the arrangement of elements and that the elements associated with each adaptor do not need to have the same composition (i.e. the 3' end may include the sequence composition of a first UID element and the 5' end may include a UID elements with different sequence composition). It will further be appreciated that some embodiments will not necessarily include the same composition of elements in each adaptor (i.e. an adaptor at the 5' end may include a UID 210 element and the adaptor on the 3' may not, or vice versa). Also, there may be inherent internal controls of the sequence quality of primer element 205 with respect to resistance to introduced error. For instance, error introduced into the sequence composition of primer 205 would negatively affect its hybridization qualities to its respective target and thus not be amplified in a PCR process and therefore not represented in populations of template molecule for sequencing. This inherent quality control of primer 205 is useful for finding UID 210, because the sequence composition of primer 205 is known and can be assumed to be substantially free of error with the exception of some sequencing related error. Also as described above, key element 207 is employed for quality control purposes and it also useful as a positional reference in the same context. Thus, in the present example primer 205 and/or key 207 may serve as easily identifiable anchor points of reference for identifying UID 210 using the known positional relationships between elements. For instance, a user or algorithm, such as an algorithm implemented by application 135, may look for UID 210 located immediately adjacent to key 207, or some known distance away, based, at least in part, upon the assumptions.

Furthermore, once a user or algorithm has identified the sequence composition of a putative UID 210 element, the step of error identification and correction occurs. Embodiments of the presently described invention compare the sequence composition of the putative UID 210 element against the sequence compositions of the UID 210 members in the set. A perfect match is associated with its sample of origin. If no perfect match is found, then the closest UID 210 elements having a sequence composition to the putative sequence are analyzed to determine possible insertion, deletion, or substitution errors that could have occurred. For example, the closest UID 210 element to the putative UID 210 element is

IDT Exhibit 1019.0214

identified or the putative UID 210 element is deemed to have too many errors. In the present example, the minimum edit distance may be computed between sequence composition of the putative UID 210 element against the sequence composition of all members of the UID 210 set or select members. The minimum edit distance may be computed using the parameters of detecting up to 3 sequence position errors with the possibility of correcting up to 2 sequence position errors. In the present example, the UID 210 member with the closest or shortest minimum edit distance to the putative UID 210 element given the parameter constraints (i.e. detection/correction) may be assigned as the sequence composition of the putative UID 210 element. Also, if the minimum edit distance calculation determines that 3 sequence position errors have occurred then, the putative UID 210 element may be assigned as unusable and not associated with a sample of origin.

Those of ordinary skill in the art will appreciate that when the UID 210 elements are employed in a combinatorial manner, each UID 210 element is typically independently analyzed. Then the combination of identified UID 210 elements may be compared against the known combinations assigned to samples of origin to identify the association of the sequence data and its specific sample of origin.

In preferred embodiments, a UID 210 finding algorithm is implemented using application 135 stored for execution on computer 130 as described above. Further, the same or other application may perform the step of associating the identified UID 210 from sequence data with the sample of origin and providing the results to a user via an interface and/or storing the results in electronic media for subsequent analysis or use.

Example 1- Design of UID elements considering a limited number of design constraints

The design of sequence composition for potential UID elements were computed considering detection, correction, and hairpin design constraints.

First a sequence length of 10 base pairs for each UID element were computed yielding 1,048,576 possible elements.

Next, of those possible elements UID elements were selected that have no monomer repeats, require only 5 flow cycles (20 flows) or less, do not begin with the "G" nucleotide species were computed yielding 34,001 possible elements.

A further step of filtering to exclude hairpins at a temperature of 40°C with a ΔG=-1.5 yielded 26,278 possible elements.

Finally, 5,000 of those possible elements were selected randomly to search for compatible sets or clusters that could correct 2 sequence position errors and detect 3 sequence position errors, yielding:

32,999 sets of 12 members

3,625 sets of 13 members

24 sets of 14 members

Example 2- Exemplary computer code for creating UID sequence elements

UIDCreate.java class file that runs a search using 1 of 3 techniques, comprising (1) based on error clouds, (2) based on edit distance, and (3) based on edit distance, with an additional efficiency strategy of using a "safety map" to precompute the edit distance which gives the software the ability to effectively look ahead in the search in advance of trying candidate selections.

```
package com.fourfivefour.amplicons;


import java.util.ArrayList;
import java.util.Arrays;
import java.util.Comparator;
import java.util.HashMap;
import java.util.HashSet;
import java.util.Iterator;
import java.util.List;
import java.util.Map;
import java.util.Set;

public class UIDCreate {
        static int maxReportSize = 0;

        public enum SearchType { ErrorCloud, EditDistance, SafetyMapEditDistance; }

        /**
         * Driver to search for a compatible set of UID sequences of a given set size
         * using one of three different search methods.
         *
         * @param candidateSequences  A set of sequences from which to select.
         * @param setSize             The desired number of sequences in the
         *                                                      solution.
         * @param errsToCorrect               The number of errors that must be able to be
         *                    corrected to obtain the original sequence.
         * @param errsToDetect        The number of errors that must be able to be
         *                    detected, but not necessarily corrected.
         * @param searchType                  The search strategy to use.
         * @return                                    A set of sequences of the requested set
size
```

```
         *                    that is a subset of the candidate sequences
         *                    and which can be distinguished according to
         *                    the error correction criteria.
         */
        public static Set<Sequence> searchCompatibleSet(Set<Sequence>
candidateSequences,
                        int setSize, int errsToCorrect, int errsToDetect, SearchType
searchType) {

                if (errsToCorrect > errsToDetect) {
                        throw new RuntimeException("The numbers of errors to correct
("+errsToCorrect+
                                                        ") must be <= the
number to detect ("+errsToDetect+")");
                }

                List<Sequence> candidates = new
ArrayList<Sequence>(candidateSequences);

                switch (searchType) {
                        case ErrorCloud:
                                return searchCompatibleSetUsingErrorClouds(candidates,
                                                setSize, errsToCorrect, errsToDetect, new
HashMap<Sequence,List<Set<Sequence>>>());
                        case EditDistance:
                                return searchCompatibleSetUsingEditDistance(candidates,
                                                setSize, errsToCorrect, errsToDetect, new
HashSet<Sequence>()
                                                );
                        case SafetyMapEditDistance:
                                return
searchCompatibleSetUsingSafetyMap(candidateSequences,

                 setSize, errsToCorrect, errsToDetect);
                        default:
                                return null;
                }
        }


        /**
         * Return a string representing the current search state.
         *
         * The resulting string encodes the current subset of sequences currently
         * under consideration, independent of the order in which they are
         * being considered.  By keeping track of this string one may ensure
         * that equivalent regions of the search space that just happen to
         * encounter sequences in a different order are not redundantly searched
         *
         * @param currentState  A set of sequences defining the current
```

```
         *                search state.
         * @return          A string that is unique for a set of sequences
         *                and independent of their order.
         */
        public static String searchStateString (Set<Sequence> currentState) {
                Sequence [] currentStateArray = currentState.toArray(new Sequence[0]);
                Arrays.sort(currentStateArray);
                StringBuffer sb = new StringBuffer();
                for (Sequence state : currentStateArray) {
                        sb.append(state.getSequence());
                        sb.append(';');
                }
                return sb.toString();
        }


        /**
         * From the set of candidates, select a set of compatible sequences that
         * can correct and detect the desired number of errors.
         *
         * This strategy uses a pre-calculated edit distance between pairs of
         * sequences to know, in advance, which sequences are compatible with
         * each other.  This pair-wise information forms a "safety map" of
         * compatible pairs that allow more efficient forward-looking knowledge
         * of what sequences will still be compatible after other sequences
         * are added to the solution set.
         *
         * @param candidateSequences Sequence set from which to select.
         * @param setSize          The desired set size to generate.
         * @param errsToCorrect     The numbers of errors that must be able
         *                        to be corrected.
         * @param errsToDetect      The number of errors that must merely be
         *                        detected.
         *
         * @return A set of sequences that meet the desired criteria.
         */
        public static Set<Sequence> searchCompatibleSetUsingSafetyMap (
                        Set<Sequence> candidateSequences,
                        int setSize, int errsToCorrect, int errsToDetect)
        {
                Map<Sequence,Set<Sequence>> safetyMap = new
HashMap<Sequence,Set<Sequence>>();

                /* Safety map is computed on a subset of the candidates in
                 * case there are too many candidates.
                 */
                final int maxSubsetSize = 3000;

                Sequence[] fullSequenceList = candidateSequences.toArray(new
Sequence[0]);
```

```
            final int numSequencesToUse =
Math.min(maxSubsetSize,candidateSequences.size());

            /* Compute the safety map, possibly using a subset of the sequences.
             * An alternative/preferential implementation might choose a random subset
             * of the sequences rather than just the fist numSequencesToUse that are
             * encountered.
             */

            for (int i=0; i < numSequencesToUse; i++) {
                Set<Sequence> iSafety = new HashSet<Sequence>();
                iSafety.add(fullSequenceList[i]);
                safetyMap.put(fullSequenceList[i], iSafety);
            }

            /* Create the safety map, ensuring that there is enough distance
             * between every pair of sequences such that sequences as far away
             * as the distance of error detection from one sequence and error
             * correction from the other sequence could not possibly overlap
             */
            int safeDistance = errsToCorrect + errsToDetect + 1;
            for (int i=0; i < numSequencesToUse; i++) {
                Set<Sequence> iSafety = safetyMap.get(fullSequenceList[i]);
                for (int j=i+1; j < numSequencesToUse; j++) {
                    Set<Sequence> jSafety = safetyMap.get(fullSequenceList[j]);

                    if (Library.minEditDistance(fullSequenceList[i],

        fullSequenceList[j])
                                                                    >=
safeDistance) {
                        iSafety.add(fullSequenceList[j]);
                        jSafety.add(fullSequenceList[i]);
                    }
                }
            }

            // Perform that actual search with the safety-map in hand
            return searchCompatibleSetUsingSafetyMap(safetyMap.keySet(),setSize,
                    errsToCorrect,errsToDetect,
                    safetyMap,new HashSet<Sequence>(), new
HashSet<String>());
        }

        /**
         * From the set of candidates, select a set of compatible sequences that
         * can correct and detect the desired number of errors.
         *
         * This strategy uses a pre-calculated edit distance between pairs of
         * sequences to know, in advance, which sequences are compatible with
```

```
     * each other.  This pair-wise information forms a "safety map" of
     * compatible pairs that allow more efficient forward-looking knowledge
     * of what sequences will still be compatible after other sequences
     * are added to the solution set.
     *
     * @param candidateSequences Sequence set from which to select.
     * @param setSize          The desired set size to generate.
     * @param errsToCorrect    The numbers of errors that must be able
     *                         to be corrected.
     * @param errsToDetect     The number of errors that must merely be
     *                         detected.
     * @param safetyMap        The map indicating the set of sequences
     *                         that are compatible with any given other
     *                         sequence.
     * @param selectedSequences  The current set of sequences selected in
     *                         the partial solution so far.
     * @param searchStateHistory A history of what subsets of sequences have
     *                         already been considered.
     *
     * @return A set of sequences that meet the desired criteria.
     */
    public static Set<Sequence> searchCompatibleSetUsingSafetyMap (
                    Set<Sequence> candidateSequences, int setSize,
                    int errsToCorrect, int errsToDetect,
                    Map<Sequence,Set<Sequence>> safetyMap,
                    Set<Sequence> selectedSequences,
                    Set<String> searchStateHistory) {


        String searchStateString = searchStateString(selectedSequences);
        if (searchStateHistory.contains(searchStateString)) {
                /* some other branch of the search has already explored this area
                 * so no need to search twice
                 */
                return null;
        } else {
                /*
                 * Record that we are initiating a search in the given state
                 */
                searchStateHistory.add(searchStateString);
        }

        if (selectedSequences.size() > maxReportSize) {
                // Provide intermediate reporting to show progress
                maxReportSize = selectedSequences.size();
                reportSet(selectedSequences);
        }
        if (selectedSequences.size() >= setSize) {
                return selectedSequences;
        }
```

```
                if (selectedSequences.size() + candidateSequences.size() < setSize) {
                        // If there aren't enough candidates to fill out the required set size,
                        // return immediately with failure.
                        return null;
                }

                /* find the candidate Sequence with the greatest number of compatible choices
of the
                 * Sequences still available:
                 */
                final Map<Sequence,Integer> remainingSafetyMapSize = new
HashMap<Sequence,Integer>();
                Sequence candidates[] = candidateSequences.toArray(new Sequence[0]);
                for (Sequence sequence : candidates) {
                        Set<Sequence> currentSafety = new
HashSet<Sequence>(safetyMap.get(sequence));
                        currentSafety.retainAll(candidateSequences);
                        remainingSafetyMapSize.put(sequence,currentSafety.size());
                }

                Comparator<Sequence> candidateSorter = new Comparator<Sequence>() {
                        public int compare(Sequence seq1, Sequence seq2) {
                                int seq1Val = remainingSafetyMapSize.get(seq1);
                                int seq2Val = remainingSafetyMapSize.get(seq2);
                                if (seq1Val > seq2Val) {
                                        return -1;
                                } else if (seq1Val < seq2Val) {
                                        return 1;
                                } else {
                                        return (seq1).compareTo(seq2);
                                }
                        }
                };

                Arrays.sort(candidates,candidateSorter);

                Set<Sequence> result = null;
                final int searchBreadth = 10;
                for (int i=0; i < candidates.length && i < searchBreadth && result == null;
i++) {
                        Sequence nextCandidate = candidates[i];
                        int numNextCompatible =
remainingSafetyMapSize.get(nextCandidate);
                        if (numNextCompatible == 1) {
                                /* the next candidate is only compatible with itself, so
                                 * the search is done
                                 */
                                break;
                        } else {
```


IDT Exhibit 1019.0221

```
                                Set<Sequence> nextCandidateSequences =
                                        new HashSet<Sequence>(candidateSequences);

        nextCandidateSequences.retainAll(safetyMap.get(nextCandidate));
                                nextCandidateSequences.remove(nextCandidate);
                                Set<Sequence> nextSelectedSeqs = new
HashSet<Sequence>(selectedSequences);
                                nextSelectedSeqs.add(nextCandidate);
                                result =

        searchCompatibleSetUsingSafetyMap(nextCandidateSequences,

         setSize,

         errsToCorrect,

         errsToDetect,

         safetyMap,

         nextSelectedSeqs,

         searchStateHistory);
                        }
                }
                return result;
        }


        /**
         * Generate the "error cloud" sequence sets out to the indicated number
         * of mutations of the given input sequence.
         *
         * This routine implements a breadth-first enumeration of all possible
         * mutations of a given sequence, while avoiding redundant, equivalent
         * subtrees of the search space.
         *
         * @param inputSequence
         *          The sequence to mutate.
         * @param numMutations
         *          The number of plys of mutations to generate
         * @return A list of sets of mutated sequences such that the i'th list
         *       element represents the set of mutated sequences that may
         *       derived via i mutations from the initial sequence and, further,
         *       cannot be derived by fewer than i mutations of the original
         *       sequence.
         */
        public static List<Set<Sequence>> generateErrorClouds(
                        Sequence inputSequence, int numMutations) {
```

```
        List<Set<Sequence>> mutationPlys = new ArrayList<Set<Sequence>>();

        // The 0'th mutation ply is simply the initial sequence itself:
        Set<Sequence> initialPly = new HashSet<Sequence>();
        initialPly.add(inputSequence);
        mutationPlys.add(initialPly);

        for (int ply = 1; ply <= numMutations; ply++) {
            Set<Sequence> lastPlySequences =
                mutationPlys.get(ply - 1);
            Set<Sequence> nextPlyCandidates =
                Sequence.generateSingleMutations(lastPlySequences);
            Iterator<Sequence> candidateIter =
                nextPlyCandidates.iterator();
            /* Remove any candidates that appeared in a previous ply
             * (meaning that they are derivable via a shorter,
             *  more direct sequence of mutations).
             */
            while (candidateIter.hasNext()) {
                Sequence candidate = candidateIter.next();
                for (Set<Sequence> previousPly : mutationPlys) {
                    if (previousPly.contains(candidate)) {
                        candidateIter.remove();
                        break;
                    }
                }
            }

            // Candidates that remain at this point represent the next ply
            mutationPlys.add(nextPlyCandidates);
        }
        return mutationPlys;
    }


    /**
     * Validate the given collection of sequences to ensure that they
     * are compatible with respect to the desired number of errors
     * to correct and detect.
     *
     * Two sequences are compatible if there is no intersection of the
     * error cloud out to the correction distance of either sequence with
     * the error cloud out to the detection distance of the other sequence.
     * If all pairs of sequences in a collection of sequences are compatible,
     * then the collection itself is compatible.
     *
     * This definition of compatibility logically follows from the fact
     * that incompatibilities exist when a potentially observable sequence
     * could be derived, via mutation, from more than one sequence
     * in the collection (thereby preventing a definitive correction to be
```

```
         * performed) and if that observable sequence is within the desired
         * error correction distance of any of those sequences.  Note that
         * mutated sequences that are within the error detection distance of
         * two sequences are not a problem as there is no expectation of being
         * able to correct such mutated sequences (only to detect that a
         * mutation has occurred).
         *
         * @param sequences          The collection of sequences to test.
         * @param errsToCorrect The desired number of errors to correct.
         * @param errsToDetect  The desired number of errors to detect.
         * @return true, if the collection of sequences are compatible,
         *       false otherwise.
         */
        public static boolean validateCollection(Set<Sequence> sequences,
                                                                    int
errsToCorrect,
                                                                    int
errsToDetect)
        {
                /* First generate the error clouds for each sequence and
                 * collapse the separate cloud levels into two sets: those
                 * sequences within the "errsToCorrect" distance of the
                 * sequence center, and those within the the "errsToDetect"
                 * distance from the center.
                 */
                Map<Sequence,Set<Sequence>> correctLevelMap =
                        new HashMap<Sequence,Set<Sequence>>();
                Map<Sequence,Set<Sequence>> detectLevelMap =
                        new HashMap<Sequence,Set<Sequence>>();
                for (Sequence sequence: sequences) {
                        List<Set<Sequence>> errorClouds =
                                generateErrorClouds(sequence, errsToDetect);
                        Set<Sequence> errsToCorrectSequences = new HashSet<Sequence>();
                        for (int errLevel=0; errLevel <= errsToCorrect; errLevel++) {
                                errsToCorrectSequences.addAll(errorClouds.get(errLevel));
                        }
                        Set<Sequence> errsToDetectSequences =
                                new HashSet<Sequence>(errsToCorrectSequences);
                        for (int errLevel=errsToCorrect+1; errLevel <= errsToDetect;
                                 errLevel++) {
                                errsToDetectSequences.addAll(errorClouds.get(errLevel));
                        }
                        correctLevelMap.put(sequence,errsToCorrectSequences);
                        detectLevelMap.put(sequence,errsToDetectSequences);
                }

                /* Validate that none of the mutations out to the "errsToDetect"
                 * error distance of one sequence are in common with any of the
                 * mutations out to the "errsToCorrect" error distance another
                 * sequence
```

```
                */
                Sequence seqList[] = sequences.toArray(new Sequence[0]);
                boolean foundOverlap = false;

                for (int i=0; !foundOverlap && i < seqList.length; i++) {
                        for (int j=0; !foundOverlap && j < seqList.length; j++) {
                                if (i == j) {
                                        continue;
                                }
                                Set<Sequence> overlapSet =
                                        new
HashSet<Sequence>(detectLevelMap.get(seqList[i]));
                                overlapSet.retainAll(correctLevelMap.get(seqList[j]));
                                foundOverlap = !overlapSet.isEmpty();
                        }
                }

                return !foundOverlap;
        }

        /**
         * Validate the given collection of sequences to ensure that they
         * are compatible with respect to the desired number of errors
         * to correct and detect.
         *
         * Two sequences are compatible if there is no intersection of the
         * error cloud out to the correction distance of either sequence with
         * the error cloud out to the detection distance of the other sequence.
         * If all pairs of sequences in a collection of sequences are compatible,
         * then the collection itself is compatible.
         *
         * This definition of compatibility logically follows from the fact
         * that incompatibilities exist when a potentially observable sequence
         * could be derived, via mutation, from more than one sequence
         * in the collection (thereby preventing a definitive correction to be
         * performed) and if that observable sequence is within the desired
         * error correction distance of any of those sequences.  Note that
         * mutated sequences that are within the error detection distance of
         * two sequences are not a problem as there is no expectation of being
         * able to correct such mutated sequences (only to detect that a
         * mutation has occurred).
         *
         * This routine performs the calculation without creating the
         * actual error clouds, but instead uses the minimum edit distance
         * between sequences to ensure compatibility.
         *
         * @param sequences         The collection of sequences to test.
         * @param errsToCorrect The desired number of errors to correct.
         * @param errsToDetect  The desired number of errors to detect.
         * @return true, if the collection of sequences are compatible,
```

```
         *       false otherwise.
         */
        public static boolean validateCollectionUsingDistance(
                                                    Set<Sequence> sequences,
                                                    int errsToCorrect, int errsToDetect) {
                Sequence seqList[] = sequences.toArray(new Sequence[0]);
                /* The error detection cloud and error correction cloud of two
                 * sequences cannot overlap if the minimum distance between the
                 * sequences is one greater than the sum of the errors.
                 */
                int safeDistance = errsToCorrect + errsToDetect + 1;
                for (int i=0; i < seqList.length; i++) {
                        for (int j=i+1; j < seqList.length; j++) {
                                int ijDist= Library.minEditDistance(seqList[i],seqList[j]);
                                if (ijDist < safeDistance) {
                                        System.out.println("Validation failed, dist = "+ijDist+"
for "+seqList[i]+" vs "+seqList[j]);
                                        return false;
                                }
                        }
                }
                return true;
        }


        /**
         * From the set of candidates, select a set of compatible sequences that
         * can correct and detect the desired number of errors.
         *
         * This strategy uses the edit distance between pairs of sequences
         * to ensure that they do not overlap in a manner that would prevent
         * error detection/correction. This technique is much more efficient
         * than explicitly enumerating and comparing the error clouds between
         * sequences.
         *
         * @param candidateSequences Sequence set from which to select.
         * @param setSize           The desired set size to generate.
         * @param errsToCorrect     The numbers of errors that must be able
         *                  to be corrected.
         * @param errsToDetect      The number of errors that must merely be
         *                  detected.
         * @param currentSequences  State of search indicating the currently
         *                  selected candidates.
         *
         * @return A set of sequences that meet the desired criteria.
         */
        public static Set<Sequence> searchCompatibleSetUsingEditDistance (
                        List<Sequence> candidateSequences, int setSize,
                        int errsToCorrect, int errsToDetect,
                        Set<Sequence> currentSequences) {
```

```
            if (currentSequences.size() > maxReportSize) {
                /**
                 * Provide intermediate level reporting to ensure progress
                 * is being made.
                 */
                maxReportSize = currentSequences.size();
                reportSet(currentSequences);
            }
            if (currentSequences.size() >= setSize) {
                return currentSequences;
            }

            /* The error detection cloud and error correction cloud of two
             * sequences cannot overlap if the minimum distance between the
             * sequences is one greater than the sum of the errors.
             */
            int safeDistance = errsToCorrect + errsToDetect + 1;

            for (int nextIdx = 0; nextIdx < candidateSequences.size(); nextIdx++) {
                /* Check to see that nextSequence is far enough away from all the
                 * other sequences in the prefix solution.
                 */
                Sequence nextSequence = candidateSequences.get(nextIdx);
                boolean overlappingClouds = false;
                for (Sequence currentSequence : currentSequences) {
                    if (Library.minEditDistance(currentSequence,nextSequence) <
safeDistance) {
                        overlappingClouds = true;
                        break;
                    }
                }

                if (!overlappingClouds) {
                    List<Sequence> remainingCandidates =
                        candidateSequences.subList(nextIdx+1,
candidateSequences.size());
                    currentSequences.add(nextSequence);

                    Set<Sequence> solution =
                        searchCompatibleSetUsingEditDistance(
                            remainingCandidates,
                            setSize, errsToCorrect, errsToDetect,
                            currentSequences);
                    if (solution == null) {
                        currentSequences.remove(nextSequence);
                    } else {
                        return solution;
                    }
                }
```


IDT Exhibit 1019.0227

```
            }
            return null;
        }

        /**
         * From the set of candidates, select a set of compatible sequences that
         * can correct and detect the desired number of errors.
         *
         * This strategy uses an explicit enumeration of the "error clouds"
         * surrounding the candidate sequences to ensure that they do not
         * overlap in a manner that would prevent error detection/correction.
         * This technique may not be appropriate, efficient, or even practically
         * usable for large error clouds.
         *
         * @param candidateSequences Sequence set from which to select.
         * @param setSize          The desired set size to generate.
         * @param errsToCorrect    The numbers of errors that must be able
         *                         to be corrected.
         * @param errsToDetect     The number of errors that must merely be
         *                         detected.
         * @param currentErrorClouds State of search mapping from the current
         *                         selected candidates and their corresponding
         *                         error clouds.
         * @return A set of sequences that meet the desired criteria.
         */
        public static Set<Sequence> searchCompatibleSetUsingErrorClouds(List<Sequence>
candidateSequences,
                                                                        int
setSize,
                                                                        int
errsToCorrect, int errsToDetect,

Map<Sequence, List<Set<Sequence>>> currentErrorClouds) {

            if (currentErrorClouds.size() > maxReportSize) {
                /**
                 * Provide intermediate level reporting to ensure progress
                 * is being made.
                 */
                maxReportSize = currentErrorClouds.size();
                reportSet(currentErrorClouds.keySet());
            }
            if (currentErrorClouds.size() >= setSize) {
                return currentErrorClouds.keySet();
            }

            for (int nextIdx = 0; nextIdx < candidateSequences.size(); nextIdx++) {
                Sequence nextSequence = candidateSequences.get(nextIdx);
                List<Set<Sequence>> nextErrorCloud =
                        generateErrorClouds(nextSequence, errsToDetect);
```

```
                    /*
                     * Check to see that nextErrorCloud doesn't overlap with any of the
                     * currentErrorClouds except possibly in the errsToDetect category,
                     * where it can only overlap with other errsToDetect distant clouds
                     */
                    boolean overlappingClouds = false;
                    for (List<Set<Sequence>> currentClouds :
currentErrorClouds.values()) {
                        for (int currentErrorLevel = 0; !overlappingClouds
                                && currentErrorLevel <= errsToCorrect;
currentErrorLevel++) {
                            for (int nextErrorLevel = 0; !overlappingClouds
                                    && nextErrorLevel <= errsToDetect;
nextErrorLevel++) {
                                Set<Sequence> currentLevelErrorCloud =
currentClouds.get(currentErrorLevel);
                                for (Sequence nextErrorSequence :
nextErrorCloud.get(nextErrorLevel)) {
                                    if
(currentLevelErrorCloud.contains(nextErrorSequence)) {
                                        overlappingClouds = true;
                                        break;
                                    }
                                }
                            }
                        }
                    }
                    /*
                     * Check to see that none of the error clouds surrounding the current
                     * codes in the range of errors to merely detect overlap with the
                     * nextErrorCloud clouds in the range of errors to correct
                     */
                    if (!overlappingClouds) {
                        for (List<Set<Sequence>> currentClouds :
currentErrorClouds.values()) {
                            for (int currentErrorLevel = errsToCorrect+1;
!overlappingClouds
                                    && currentErrorLevel <= errsToDetect;
currentErrorLevel++) {
                                for (int nextErrorLevel = 0; !overlappingClouds
                                        && nextErrorLevel <=
errsToCorrect; nextErrorLevel++) {
                                    Set<Sequence> currentLevelErrorCloud
= currentClouds.get(currentErrorLevel);
                                    for (Sequence nextErrorSequence :
nextErrorCloud.get(nextErrorLevel)) {
                                        if
(currentLevelErrorCloud.contains(nextErrorSequence)) {
                                            overlappingClouds = true;
```


IDT Exhibit 1019.0229

```
                                                    break;
                                                }
                                            }
                                        }
                                    }
                                }
                            }

                            if (!overlappingClouds) {
                                List<Sequence> remainingCandidates =
                                        candidateSequences.subList(nextIdx+1,
candidateSequences.size());
                                currentErrorClouds.put(nextSequence, nextErrorCloud);

                                Set<Sequence> solution =
searchCompatibleSetUsingErrorClouds(
                                                    remainingCandidates,
                                                    setSize, errsToCorrect, errsToDetect,
currentErrorClouds);
                                if (solution == null) {
                                        currentErrorClouds.remove(nextSequence);
                                } else {
                                        return solution;
                                }
                            }
                        }
                        return null;
                }

                /**
                 * Generate the set of all possible flow patterns that would produce
                 * sequences of a given length, with no monomer repeats, using at most
                 * a given number of flows, and avoiding a given starting flow as the
                 * first positive flow of the pattern.
                 *
                 * @param numBases      The number of called bases that should
                 *                      be produced if the given flowgram were
                 *                      observed for a sequence.
                 * @param startFlow     The conceptual flow index of the first
                 *                      flow in the recursive call to this routine.
                 * @param maxFlow       The maximum number of flows (positive or
                 *                      negative) that may be consumed by the pattern.
                 * @param avoidStartFlow  A flow that may not be the first positive flow.
                 * @return Set of flow patterns (Strings of 0's and 1's representing
                 *      negative and positive flows that meet the desired
                 *      criteria).
                 */
                public static Set<String> generateCandidateFlowPatterns(int numBases,
                                int startFlow, int maxFlow, int avoidStartFlow) {
                        Set<String> candidates = new HashSet<String>();
```

```
                if (numBases > 0) {
                        if (startFlow == maxFlow) {
                                return candidates;
                        } else {
                                String zeroPrefix = "";
                                for (int flowSkip = 0; flowSkip < (startFlow == 0 ? 4 : 3);
flowSkip++) {
                                        int nextPositiveFlow = startFlow + flowSkip;
                                        if (nextPositiveFlow < maxFlow
                                                        && (startFlow > 0 || nextPositiveFlow !=
avoidStartFlow)) {
                                                String currentPrefix = zeroPrefix + "1";
                                                for (String candidateSuffix :
generateCandidateFlowPatterns(
                                                                numBases - 1, nextPositiveFlow
+ 1, maxFlow,
                                                                avoidStartFlow)) {
                                                        candidates.add(currentPrefix +
candidateSuffix);
                                                }
                                        }
                                        zeroPrefix += "0";
                                }
                        }
                } else {
                        candidates.add("");
                }
                return candidates;
        }

        /**
         * Generate a set of candidate sequences of a given base content length, using
         * a maximum number of flow cycles, based on the given flow order such that
         * none of the sequences begin with a particular base.
         *
         * @param numBases   Number of bases each candidate sequence should be.
         * @param maxCycles  Maximum number of flow cycles allowed.
         * @param flowOrder  Flow order for sequencing.
         * @param avoidStartBase Base to avoid at start of sequence.
         * @return Candidate set of sequences that meet the given criteria.
         */
        public static Set<Sequence> generateCandidateSequences(int numBases, int
maxCycles,
                        String flowOrder, char avoidStartBase)
        {
                Set<String> candidateFlows =
                        generateCandidateFlowPatterns(numBases,
                                                        0,
                                                        maxCycles*4,
```


IDT Exhibit 1019.0231

```
flowOrder.indexOf(avoidStartBase));

                Set<Sequence> candidateSequences = new HashSet<Sequence>();
                for (String flows : candidateFlows) {
                        Flowgram f = new
Flowgram(flowOrder,Flowgram.flowValueStringToFlowValues(flows));
                        candidateSequences.add(new Sequence(f.baseCall()));
                }

                /*
                 * At this point, one might further filter the candidate sequences by some
                 * molecular biological criteria, such as propensity to form hairpin turns
                 */
                return candidateSequences;
        }


        /**
         * Produce a printed representation of a set of sequences
         * assuming a default flow oder of TACG.
         *
         * @param sequences  The set of sequences to report.
         */
        public static void reportSet(Set<Sequence> sequences) {
                reportSet(sequences, "TACG");
        }

        /**
         * Produce a printed representation of a set of sequences
         * along with their conceptual flowgrams, assuming the
         * given flow order.
         *
         * @param sequenceSet The set of sequences to report.
         * @param flowOrder   The flow order that would be used
         *                    in sequencing.
         */

        public static void reportSet(Set<Sequence> sequenceSet,String flowOrder) {
                Sequence sequences[] = sequenceSet.toArray(new Sequence[0]);
                Arrays.sort(sequences);

                for (int i = 0; i < sequences.length; i++) {
                        System.out.println("Sequence #" + (i + 1) + "\t"
                                        + new
Flowgram(flowOrder,sequences[i].getSequence()).getFlowValueString() + "\t"
                                        + sequences[i].getSequence());
                }
                System.out.println("---------");
                System.out.flush();
```


IDT Exhibit 1019.0232

```
        }

        /**
         * Example use of search code to generate a set of compatible
         * sequences that meet given design criteria of base composition,
         * maximum flow usage, and ability to correct and detect errors.
         *
         * @param args
         */
        public static void main(String[] args) {
                char fivePrimeAvoidBase    = 'G';
                final String flowOrder     = "TACG";
                final int   uidBaseLength  = 10;
                final int   uidMaxCycles   = 5;
                final int   errorsToCorrect = 2;
                final int   errorsToDetect  = 3;
                final int   desiredSetSize  = 12;
                System.out.println("Computation Begun");

                Set<Sequence> sequenceSet =
                        generateCandidateSequences(uidBaseLength,
                                                   uidMaxCycles,
                                                   flowOrder,
                                                   fivePrimeAvoidBase);
                System.out.println("Total candidate sequencess=" + sequenceSet.size());

                Set<Sequence> compatibleSequences =
                        searchCompatibleSet(sequenceSet,desiredSetSize,
                                            errorsToCorrect,errorsToDetect,
                                            SearchType.EditDistance);
                if (compatibleSequences == null) {
                        System.out.println("No solution");
                } else {
                        reportSet(compatibleSequences,flowOrder);
                }
        }
}
```

It will be appreciated that the foregoing computer code is provided for the purposes of example, and that numerous alternative methods and code structures may be employed. It will also be appreciated that the exemplary code provided herein is not intended to execute as a stand alone application or to run perfectly without additional computer code or modification.

Example 3- Table of computed UID sequences, cluster ID, and Flowgram script

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127176	14	01100101010110011010	ACAGAGTGTC	10	7
C1127176	14	01111010100101010100	ACGTCTGAGA	10	8
C1127176	14	01010111001001101010	AGACGCACTC	10	9
C1127176	14	01001010110010101011	ATCTATCTCG	10	10
C1127176	14	00110100111100111000	CGATACGCGT	10	11
C1127176	14	00110011001110010011	CGCGCGTGCG	10	12
C1127176	14	00111101010011010010	CGTAGATAGC	10	13
C1127176	14	00111001101010101100	CGTGTCTCTA	10	14
C1127176	14	00101010011001110110	CTCACACGAC	10	15
C1127176	14	11101010010010111000	TACTCATCGT	10	16
C1127176	14	11010011010011100100	TAGCGATACA	10	17
C1127176	14	11001001110111001000	TATGTAGTAT	10	18
C1127176	14	10101001001101101001	TCTGCGACTG	10	19
C1127176	14	10010110010110100101	TGACAGTCAG	10	20
C1127177	14	01101101001101010100	ACTAGCGAGA	10	21
C1127177	14	01010111010011001100	AGACGATATA	10	22
C1127177	14	01001010100101111010	ATCTGACGTC	10	23
C1127177	14	01001001101011010011	ATGTCTAGCG	10	24
C1127177	14	00110100111100111000	CGATACGCGT	10	25
C1127177	14	00110011001110010011	CGCGCGTGCG	10	26
C1127177	14	00111010011001010110	CGTCACAGAC	10	27
C1127177	14	00111001101010101100	CGTGTCTCTA	10	28
C1127177	14	11101010010101001001	TACTCAGATG	10	29
C1127177	14	11010010011010101010	TAGCACTCTC	10	30
C1127177	14	11001100111001100100	TATATACACA	10	31
C1127177	14	10100100101110100101	TCATCGTCAG	10	32
C1127177	14	10010101100100110110	TGAGTGCGAC	10	33
C1127177	14	10011001010111011000	TGTGAGTAGT	10	34
C1127178	14	01100110101010010110	ACACTCTGAC	10	35
C1127178	14	01010101010101101001	AGAGAGACTG	10	36
C1127178	14	01001111110010101000	ATACGTATCT	10	37
C1127178	14	01001011101101010100	ATCGTCGAGA	10	38
C1127178	14	00100110010111011100	CACAGTAGTA	10	39
C1127178	14	00110100111100111000	CGATACGCGT	10	40
C1127178	14	00110011001110010011	CGCGCGTGCG	10	41
C1127178	14	00111001101010101100	CGTGTCTCTA	10	42
C1127178	14	00101001110101001011	CTGTAGATCG	10	43
C1127178	14	11101001010100110010	TACTGAGCGC	10	44
C1127178	14	11010010101111001000	TAGCTCGTAT	10	45
C1127178	14	11001100111001100100	TATATACACA	10	46
C1127178	14	10110010011001101010	TCGCACACTC	10	47
C1127178	14	10101100100110011001	TCTATGTGTG	10	48
C1127179	14	01101011011111000000	ACTCGACGTA	10	49
C1127179	14	01010110100111010100	AGACTGTAGA	10	50
C1127179	14	01010101010101101001	AGAGAGACTG	10	51

IDT Exhibit 1019.0234

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127179	14	01001001101011010011	ATGTCTAGCG	10	52
C1127179	14	00100110111011001001	CACTACTATG	10	53
C1127179	14	00110100111100111000	CGATACGCGT	10	54
C1127179	14	00110011001110010011	CGCGCGTGCG	10	55
C1127179	14	00111010011001010110	CGTCACAGAC	10	56
C1127179	14	00111001101010101100	CGTGTCTCTA	10	57
C1127179	14	11110101001001010010	TACGAGCAGC	10	58
C1127179	14	11010010010010111001	TAGCATCGTG	10	59
C1127179	14	11001110011010100100	TATACACTCA	10	60
C1127179	14	10101001100110010110	TCTGTGTGAC	10	61
C1127179	14	10011101111001001000	TGTAGTACAT	10	62
C1127180	14	01101011010010101010	ACTCGATCTC	10	63
C1127180	14	01010110100111010100	AGACTGTAGA	10	64
C1127180	14	01010101010101101001	AGAGAGACTG	10	65
C1127180	14	01001001101011010011	ATGTCTAGCG	10	66
C1127180	14	00100110111011001001	CACTACTATG	10	67
C1127180	14	00110100111100111000	CGATACGCGT	10	68
C1127180	14	00110011001110010011	CGCGCGTGCG	10	69
C1127180	14	00111010011001010110	CGTCACAGAC	10	70
C1127180	14	00111001101010101100	CGTGTCTCTA	10	71
C1127180	14	11110101001001010010	TACGAGCAGC	10	72
C1127180	14	11010010010010111001	TAGCATCGTG	10	73
C1127180	14	11001110011010100100	TATACACTCA	10	74
C1127180	14	10101001100110010110	TCTGTGTGAC	10	75
C1127180	14	10011101111001001000	TGTAGTACAT	10	76
C1127181	14	01100110011100101001	ACACACGCTG	10	77
C1127181	14	01110100101001001101	ACGATCATAG	10	78
C1127181	14	01010101010101100110	AGAGAGACAC	10	79
C1127181	14	01001110110010010110	ATACTATGAC	10	80
C1127181	14	00110011001110010011	CGCGCGTGCG	10	81
C1127181	14	00111001101010101100	CGTGTCTCTA	10	82
C1127181	14	00101111011001011000	CTACGACAGT	10	83
C1127181	14	00101001110101001011	CTGTAGATCG	10	84
C1127181	14	11010010010110101100	TAGCAGTCTA	10	85
C1127181	14	11011001001100111000	TAGTGCGCGT	10	86
C1127181	14	10101100100110011001	TCTATGTGTG	10	87
C1127181	14	10101011001010100110	TCTCGCTCAC	10	88
C1127181	14	10010100111011101000	TGATACTACT	10	89
C1127181	14	10011010110101010100	TGTCTAGAGA	10	90
C1127182	14	01100101101011110000	ACAGTCTACG	10	91
C1127182	14	01010111001001101010	AGACGCACTC	10	92
C1127182	14	01010010111001001101	AGCTACATAG	10	93
C1127182	14	01011010100110010110	AGTCTGTGAC	10	94
C1127182	14	01001101010110011100	ATAGAGTGTA	10	95
C1127182	14	00110011001110010011	CGCGCGTGCG	10	96

IDT Exhibit 1019.0235

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127182	14	00111001101010101100	CGTGTCTCTA	10	97
C1127182	14	00101110110100101001	CTACTAGCTG	10	98
C1127182	14	00101001010101110101	CTGAGACGAG	10	99
C1127182	14	11011001001100111000	TAGTGCGCGT	10	100
C1127182	14	10100111110010010100	TCACGTATGA	10	101
C1127182	14	10111010010101001010	TCGTCAGATC	10	102
C1127182	14	10101100111001100100	TCTATACACA	10	103
C1127182	14	10010100110110110010	TGATAGTCGC	10	104
C1127183	14	01110100101100111000	ACGATCGCGT	10	105
C1127183	14	01101010110010011001	ACTCTATGTG	10	106
C1127183	14	01010010011001101101	AGCACACTAG	10	107
C1127183	14	01001110010101011010	ATACAGAGTC	10	108
C1127183	14	01001100101010100111	ATATCTCACG	10	109
C1127183	14	00100101110011110010	CAGTATACGC	10	110
C1127183	14	00110011001110010011	CGCGCGTGCG	10	111
C1127183	14	00111001101010101100	CGTGTCTCTA	10	112
C1127183	14	00101111111001001000	CTACGTACAT	10	113
C1127183	14	11001111001010010100	TATACGCTGA	10	114
C1127183	14	10110110010010101010	TCGACATCTC	10	115
C1127183	14	10110010110101100100	TCGCTAGACA	10	116
C1127183	14	10010101100100110110	TGAGTGCGAC	10	117
C1127183	14	10011001010111011000	TGTGAGTAGT	10	118
C1127184	14	01100111001010100110	ACACGCTCAC	10	119
C1127184	14	01110100101100111000	ACGATCGCGT	10	120
C1127184	14	01010111010101010100	AGACGAGAGA	10	121
C1127184	14	01010010100111001110	AGCTGTATAC	10	122
C1127184	14	01001101100101001011	ATAGTGATCG	10	123
C1127184	14	00100110111001101001	CACTACACTG	10	124
C1127184	14	00110011001110010011	CGCGCGTGCG	10	125
C1127184	14	00111101011101100000	CGTAGACGAC	10	126
C1127184	14	00111001101010101100	CGTGTCTCTA	10	127
C1127184	14	11100100110010010101	TACATATGAG	10	128
C1127184	14	10101010101101100100	TCTCTCGACA	10	129
C1127184	14	10101001010100101101	TCTGAGCTAG	10	130
C1127184	14	10010101010011101010	TGAGATACTC	10	131
C1127184	14	10011110100110011000	TGTACTGTGT	10	132
C1127185	14	01100100101110101001	ACATCGTCTG	10	133
C1127185	14	01110010100111011000	ACGCTGTAGT	10	134
C1127185	14	01010101010101100110	AGAGAGACAC	10	135
C1127185	14	01011010010100111100	AGTCAGCGTA	10	136
C1127185	14	01001111001001110100	ATACGCACGA	10	137
C1127185	14	00100100100111010111	CATGTAGACG	10	138
C1127185	14	00110011001110010011	CGCGCGTGCG	10	139
C1127185	14	00111001101010101100	CGTGTCTCTA	10	140
C1127185	14	00101110010010011110	CTACATGTAC	10	141

IDT Exhibit 1019.0236

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127185	14	11101110100100101000	TACTACTGCT	10	142
C1127185	14	11010101010010011001	TAGAGATGTG	10	143
C1127185	14	10100101011011010100	TCAGACTAGA	10	144
C1127185	14	10011100101101010010	TGTATCGAGC	10	145
C1127185	14	10011011111001001000	TGTCGTACAT	10	146
C1127186	14	01100100101110101001	ACATCGTCTG	10	147
C1127186	14	01110010100111011000	ACGCTGTAGT	10	148
C1127186	14	01010101010101100110	AGAGAGACAC	10	149
C1127186	14	01011010010100111100	AGTCAGCGTA	10	150
C1127186	14	01001111001001110100	ATACGCACGA	10	151
C1127186	14	00100100100111010111	CATGTAGACG	10	152
C1127186	14	00110011001110010011	CGCGCGTGCG	10	153
C1127186	14	00111001101010101100	CGTGTCTCTA	10	154
C1127186	14	00101110010010011110	CTACATGTAC	10	155
C1127186	14	11101110100100101000	TACTACTGCT	10	156
C1127186	14	11010101010010011001	TAGAGATGTG	10	157
C1127186	14	10100101011011010100	TCAGACTAGA	10	158
C1127186	14	10110010011001101010	TCGCACACTC	10	159
C1127186	14	10011100101101010010	TGTATCGAGC	10	160
C1127187	14	01100111001010100110	ACACGCTCAC	10	161
C1127187	14	01110010100111011000	ACGCTGTAGT	10	162
C1127187	14	01011010010010111010	AGTCATCGTC	10	163
C1127187	14	01011001010101100101	AGTGAGACAG	10	164
C1127187	14	01001101010110011100	ATAGAGTGTA	10	165
C1127187	14	00100110010011110101	CACATACGAG	10	166
C1127187	14	00110011001110010011	CGCGCGTGCG	10	167
C1127187	14	00111001101010101100	CGTGTCTCTA	10	168
C1127187	14	00101010110101101010	CTCTAGACTC	10	169
C1127187	14	11001110101001010100	TATACTCAGA	10	170
C1127187	14	11001011110010110000	TATCGTATCG	10	171
C1127187	14	10111110010011001000	TCGTACATAT	10	172
C1127187	14	10101001100110010110	TCTGTGTGAC	10	173
C1127187	14	10010111011100101000	TGACGACGCT	10	174
C1127188	14	01100100101110101001	ACATCGTCTG	10	175
C1127188	14	01110010011101001100	ACGCACGATA	10	176
C1127188	14	01010101110100111000	AGAGTAGCGT	10	177
C1127188	14	01011110011001101000	AGTACACACT	10	178
C1127188	14	01011010100110010110	AGTCTGTGAC	10	179
C1127188	14	00100110111011010010	CACTACTAGC	10	180
C1127188	14	00110011001110010011	CGCGCGTGCG	10	181
C1127188	14	00111001101010101100	CGTGTCTCTA	10	182
C1127188	14	00101110010110011100	CTACAGTGTA	10	183
C1127188	14	00101001110101001011	CTGTAGATCG	10	184
C1127188	14	11001011001101011000	TATCGCGAGT	10	185
C1127188	14	10110110010100100101	TCGACAGCAG	10	186

IDT Exhibit 1019.0237

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127188	14	10101010010011110100	TCTCATACGA	10	187
C1127188	14	10010101010011001110	TGAGATATAC	10	188
C1127189	14	01100101001010110110	ACAGCTCGAC	10	189
C1127189	14	01101011010011100100	ACTCGATACA	10	190
C1127189	14	01010100110101101100	AGATAGACTA	10	191
C1127189	14	01010011001110011001	AGCGCGTGTG	10	192
C1127189	14	01001001101011010011	ATGTCTAGCG	10	193
C1127189	14	00100111110111001000	CACGTAGTAT	10	194
C1127189	14	00110110011100100101	CGACACGCAG	10	195
C1127189	14	00111001010101010110	CGTGAGAGAC	10	196
C1127189	14	00111001101010101100	CGTGTCTCTA	10	197
C1127189	14	11101100101100101000	TACTATCGCT	10	198
C1127189	14	11011001001001100101	TAGTGCACAG	10	199
C1127189	14	10110010010101111000	TCGCAGACGT	10	200
C1127189	14	10101110100110010100	TCTACTGTGA	10	201
C1127189	14	10010111101001001010	TGACGTCATC	10	202
C1127190	14	01100101011001001101	ACAGACATAG	10	203
C1127190	14	01110011001110011000	ACGCGCGTGT	10	204
C1127190	14	01010111001001101010	AGACGCACTC	10	205
C1127190	14	01010010110010110101	AGCTATCGAG	10	206
C1127190	14	01001100100110011110	ATATGTGTAC	10	207
C1127190	14	01001001110101111000	ATGTAGACGT	10	208
C1127190	14	00100110101110100110	CACTCGTCAC	10	209
C1127190	14	00111001010101010110	CGTGAGAGAC	10	210
C1127190	14	00111001101010101100	CGTGTCTCTA	10	211
C1127190	14	00101010011011011001	CTCACTAGTG	10	212
C1127190	14	11101100110010101000	TACTATATCT	10	213
C1127190	14	11001110101001010100	TATACTCAGA	10	214
C1127190	14	10110010010111100100	TCGCAGTACA	10	215
C1127190	14	10010101100100110011	TGAGTGCGCG	10	216
C1127191	14	01100101011010011010	ACAGACTGTC	10	217
C1127191	14	01010011001111100100	AGCGCGTACA	10	218
C1127191	14	01011011010100101001	AGTCGAGCTG	10	219
C1127191	14	01001110110101010010	ATACTAGAGC	10	220
C1127191	14	01001011101001111000	ATCGTCACGT	10	221
C1127191	14	00110101001001110011	CGAGCACGCG	10	222
C1127191	14	00111001010101010110	CGTGAGAGAC	10	223
C1127191	14	00111001101010101100	CGTGTCTCTA	10	224
C1127191	14	00101010010010111101	CTCATCGTAG	10	225
C1127191	14	11100110101100101000	TACACTCGCT	10	226
C1127191	14	11001001110111001000	TATGTAGTAT	10	227
C1127191	14	10110010100110011001	TCGCTGTGTG	10	228
C1127191	14	10101100111001100100	TCTATACACA	10	229
C1127191	14	10010111101001001010	TGACGTCATC	10	230
C1127192	14	01101001100101001011	ACTGTGATCG	10	231

IDT Exhibit 1019.0238

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127192	14	01010011001111100100	AGCGCGTACA	10	232
C1127192	14	01011110100110011000	AGTACTGTGT	10	233
C1127192	14	01001110101001100110	ATACTCACAC	10	234
C1127192	14	00100110010111011100	CACAGTAGTA	10	235
C1127192	14	00110101001001110011	CGAGCACGCG	10	236
C1127192	14	00111001010101010110	CGTGAGAGAC	10	237
C1127192	14	00111001101010101100	CGTGTCTCTA	10	238
C1127192	14	00101010010010101111	CTCATCTACG	10	239
C1127192	14	10110010110010010101	TCGCTATGAG	10	240
C1127192	14	10101100111001001100	TCTATACATA	10	241
C1127192	14	10101011001100111000	TCTCGCGCGT	10	242
C1127192	14	10010111011011001000	TGACGACTAT	10	243
C1127192	14	10010100110110110010	TGATAGTCGC	10	244
C1127193	14	01101001100101001011	ACTGTGATCG	10	245
C1127193	14	01010011001111100100	AGCGCGTACA	10	246
C1127193	14	01011110100110011000	AGTACTGTGT	10	247
C1127193	14	00100110010111011100	CACAGTAGTA	10	248
C1127193	14	00110101001001110011	CGAGCACGCG	10	249
C1127193	14	00111001010101010110	CGTGAGAGAC	10	250
C1127193	14	00111001101010101100	CGTGTCTCTA	10	251
C1127193	14	00101010010010101111	CTCATCTACG	10	252
C1127193	14	11010010101001101010	TAGCTCACTC	10	253
C1127193	14	10110010110010010101	TCGCTATGAG	10	254
C1127193	14	10101100111001001100	TCTATACATA	10	255
C1127193	14	10101011001100111000	TCTCGCGCGT	10	256
C1127193	14	10010111011011001000	TGACGACTAT	10	257
C1127193	14	10010100110110110010	TGATAGTCGC	10	258
C1127194	14	01101001100101001011	ACTGTGATCG	10	259
C1127194	14	01010011001111100100	AGCGCGTACA	10	260
C1127194	14	01011100111001011000	AGTATACAGT	10	261
C1127194	14	00100100111110011001	CATACGTGTG	10	262
C1127194	14	00110101001001110011	CGAGCACGCG	10	263
C1127194	14	00111001010101010110	CGTGAGAGAC	10	264
C1127194	14	00111001101010101100	CGTGTCTCTA	10	265
C1127194	14	00101010010010101111	CTCATCTACG	10	266
C1127194	14	11100100110101001100	TACATAGATA	10	267
C1127194	14	11010010101001101010	TAGCTCACTC	10	268
C1127194	14	10110010110010010101	TCGCTATGAG	10	269
C1127194	14	10101011001100111000	TCTCGCGCGT	10	270
C1127194	14	10010111011011001000	TGACGACTAT	10	271
C1127194	14	10010100110110110010	TGATAGTCGC	10	272
C1127195	14	01101110101001010100	ACTACTCAGA	10	273
C1127195	14	01101001100101001011	ACTGTGATCG	10	274
C1127195	14	01010011001111100100	AGCGCGTACA	10	275
C1127195	14	00100100111110011001	CATACGTGTG	10	276

IDT Exhibit 1019.0239

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127195	14	00110101001001110011	CGAGCACGCG	10	277
C1127195	14	00111001010101010110	CGTGAGAGAC	10	278
C1127195	14	00111001101010101100	CGTGTCTCTA	10	279
C1127195	14	00101010010010101111	CTCATCTACG	10	280
C1127195	14	11100100110101001100	TACATAGATA	10	281
C1127195	14	11010010101001101010	TAGCTCACTC	10	282
C1127195	14	10110010110010010101	TCGCTATGAG	10	283
C1127195	14	10101011001100111000	TCTCGCGCGT	10	284
C1127195	14	10010111011011001000	TGACGACTAT	10	285
C1127195	14	10010100110110110010	TGATAGTCGC	10	286
C1127196	14	01100101011010011010	ACAGACTGTC	10	287
C1127196	14	01101011001100101001	ACTCGCGCTG	10	288
C1127196	14	01010011001111100100	AGCGCGTACA	10	289
C1127196	14	01011100111001001001	AGTATACATG	10	290
C1127196	14	01001110010110110100	ATACAGTCGA	10	291
C1127196	14	00100111110111001000	CACGTAGTAT	10	292
C1127196	14	00110101001001110011	CGAGCACGCG	10	293
C1127196	14	00111001010101010110	CGTGAGAGAC	10	294
C1127196	14	00111001101010101100	CGTGTCTCTA	10	295
C1127196	14	00101010010010101111	CTCATCTACG	10	296
C1127196	14	11010010101001101010	TAGCTCACTC	10	297
C1127196	14	11001001100110010011	TATGTGTGCG	10	298
C1127196	14	10110100101110010100	TCGATCGTGA	10	299
C1127196	14	10010110100101001101	TGACTGATAG	10	300
C1127197	14	01100101011001001101	ACAGACATAG	10	301
C1127197	14	01101011001100101001	ACTCGCGCTG	10	302
C1127197	14	01010011001111100100	AGCGCGTACA	10	303
C1127197	14	01011100111001011000	AGTATACAGT	10	304
C1127197	14	01001110010110110100	ATACAGTCGA	10	305
C1127197	14	00100111110111001000	CACGTAGTAT	10	306
C1127197	14	00110101001001110011	CGAGCACGCG	10	307
C1127197	14	00111001010101010110	CGTGAGAGAC	10	308
C1127197	14	00111001101010101100	CGTGTCTCTA	10	309
C1127197	14	00101010010010101111	CTCATCTACG	10	310
C1127197	14	11010100110010101010	TAGATATCTC	10	311
C1127197	14	11001001100110010011	TATGTGTGCG	10	312
C1127197	14	10110100101110010100	TCGATCGTGA	10	313
C1127197	14	10011110100101001001	TGTACTGATG	10	314
C1127198	14	01100101011001001101	ACAGACATAG	10	315
C1127198	14	01101011001100101001	ACTCGCGCTG	10	316
C1127198	14	01010011001111100100	AGCGCGTACA	10	317
C1127198	14	01011100111001011000	AGTATACAGT	10	318
C1127198	14	01001110010110110100	ATACAGTCGA	10	319
C1127198	14	00100111110111001000	CACGTAGTAT	10	320
C1127198	14	00110101001001110011	CGAGCACGCG	10	321

IDT Exhibit 1019.0240

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127198	14	00111001010101010110	CGTGAGAGAC	10	322
C1127198	14	00111001101010101100	CGTGTCTCTA	10	323
C1127198	14	00101010010010101111	CTCATCTACG	10	324
C1127198	14	11010010110100101010	TAGCTAGCTC	10	325
C1127198	14	11001001100110010011	TATGTGTGCG	10	326
C1127198	14	10110100101110010100	TCGATCGTGA	10	327
C1127198	14	10011110100101001001	TGTACTGATG	10	328
C1127199	14	01100101011001001101	ACAGACATAG	10	329
C1127199	14	01101011001100101001	ACTCGCGCTG	10	330
C1127199	14	01010011001111100100	AGCGCGTACA	10	331
C1127199	14	01011100111001011000	AGTATACAGT	10	332
C1127199	14	01001110010110110100	ATACAGTCGA	10	333
C1127199	14	00100111110111001000	CACGTAGTAT	10	334
C1127199	14	00110101001001110011	CGAGCACGCG	10	335
C1127199	14	00111001010101010110	CGTGAGAGAC	10	336
C1127199	14	00111001101010101100	CGTGTCTCTA	10	337
C1127199	14	00101010010010101111	CTCATCTACG	10	338
C1127199	14	11010010101001101010	TAGCTCACTC	10	339
C1127199	14	11001001100110010011	TATGTGTGCG	10	340
C1127199	14	10110100101110010100	TCGATCGTGA	10	341
C1127199	14	10011110100101001001	TGTACTGATG	10	342
C1127200	14	01100101011001001101	ACAGACATAG	10	343
C1127200	14	01101011001100101001	ACTCGCGCTG	10	344
C1127200	14	01010011001111100100	AGCGCGTACA	10	345
C1127200	14	01001110010110110100	ATACAGTCGA	10	346
C1127200	14	00100111110111001000	CACGTAGTAT	10	347
C1127200	14	00110101001001110011	CGAGCACGCG	10	348
C1127200	14	00111001010101010110	CGTGAGAGAC	10	349
C1127200	14	00111001101010101100	CGTGTCTCTA	10	350
C1127200	14	00101010010010101111	CTCATCTACG	10	351
C1127200	14	11010100110010101010	TAGATATCTC	10	352
C1127200	14	11010011011001011000	TAGCGACAGT	10	353
C1127200	14	11001001100110010011	TATGTGTGCG	10	354
C1127200	14	10110100101110010100	TCGATCGTGA	10	355
C1127200	14	10011110100101001001	TGTACTGATG	10	356
C1127201	14	01101011001100101001	ACTCGCGCTG	10	357
C1127201	14	01010011001111100100	AGCGCGTACA	10	358
C1127201	14	01001110010110110100	ATACAGTCGA	10	359
C1127201	14	01001010111001100110	ATCTACACAC	10	360
C1127201	14	00100111110111001000	CACGTAGTAT	10	361
C1127201	14	00110101001001110011	CGAGCACGCG	10	362
C1127201	14	00111001010101010110	CGTGAGAGAC	10	363
C1127201	14	00111001101010101100	CGTGTCTCTA	10	364
C1127201	14	00101010010010101111	CTCATCTACG	10	365
C1127201	14	11010100110010101010	TAGATATCTC	10	366

IDT Exhibit 1019.0241

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127201	14	11010011011001011000	TAGCGACAGT	10	367
C1127201	14	11001001100110010011	TATGTGTGCG	10	368
C1127201	14	10110100101110010100	TCGATCGTGA	10	369
C1127201	14	10011110100101001001	TGTACTGATG	10	370
C1127202	14	01100101011001001101	ACAGACATAG	10	371
C1127202	14	01101011001100101001	ACTCGCGCTG	10	372
C1127202	14	01101010110010010110	ACTCTATGAC	10	373
C1127202	14	01010011001111100100	AGCGCGTACA	10	374
C1127202	14	01011100111001011000	AGTATACAGT	10	375
C1127202	14	01001110010110110100	ATACAGTCGA	10	376
C1127202	14	00100111110111001000	CACGTAGTAT	10	377
C1127202	14	00110101001001110011	CGAGCACGCG	10	378
C1127202	14	00111001010101010110	CGTGAGAGAC	10	379
C1127202	14	00111001101010101100	CGTGTCTCTA	10	380
C1127202	14	11010100110010101010	TAGATATCTC	10	381
C1127202	14	11001001100110010011	TATGTGTGCG	10	382
C1127202	14	10110100101110010100	TCGATCGTGA	10	383
C1127202	14	10011110100101001001	TGTACTGATG	10	384
C1127203	14	01100101011001001101	ACAGACATAG	10	385
C1127203	14	01101011001100101001	ACTCGCGCTG	10	386
C1127203	14	01101010110010010110	ACTCTATGAC	10	387
C1127203	14	01010011001111100100	AGCGCGTACA	10	388
C1127203	14	01011100111001011000	AGTATACAGT	10	389
C1127203	14	01001110010110110100	ATACAGTCGA	10	390
C1127203	14	00100111110111001000	CACGTAGTAT	10	391
C1127203	14	00110101001001110011	CGAGCACGCG	10	392
C1127203	14	00111001010101010110	CGTGAGAGAC	10	393
C1127203	14	00111001101010101100	CGTGTCTCTA	10	394
C1127203	14	11010010110100101010	TAGCTAGCTC	10	395
C1127203	14	11001001100110010011	TATGTGTGCG	10	396
C1127203	14	10110100101110010100	TCGATCGTGA	10	397
C1127203	14	10011110100101001001	TGTACTGATG	10	398
C1127204	14	01100101011001001101	ACAGACATAG	10	399
C1127204	14	01101011001100101001	ACTCGCGCTG	10	400
C1127204	14	01101010110010010110	ACTCTATGAC	10	401
C1127204	14	01010011001111100100	AGCGCGTACA	10	402
C1127204	14	01011100111001011000	AGTATACAGT	10	403
C1127204	14	01001110010110110100	ATACAGTCGA	10	404
C1127204	14	00100111110111001000	CACGTAGTAT	10	405
C1127204	14	00110101001001110011	CGAGCACGCG	10	406
C1127204	14	00111001010101010110	CGTGAGAGAC	10	407
C1127204	14	00111001101010101100	CGTGTCTCTA	10	408
C1127204	14	11010010101001101010	TAGCTCACTC	10	409
C1127204	14	11001001100110010011	TATGTGTGCG	10	410
C1127204	14	10110100101110010100	TCGATCGTGA	10	411

IDT Exhibit 1019.0242

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127204	14	10011110100101001001	TGTACTGATG	10	412
C1127205	14	01100101011001001101	ACAGACATAG	10	413
C1127205	14	01101011001100101001	ACTCGCGCTG	10	414
C1127205	14	01101010110010010110	ACTCTATGAC	10	415
C1127205	14	01010011001111100100	AGCGCGTACA	10	416
C1127205	14	01001110010110110100	ATACAGTCGA	10	417
C1127205	14	00100111110111001000	CACGTAGTAT	10	418
C1127205	14	00110101001001110011	CGAGCACGCG	10	419
C1127205	14	00111001010101010110	CGTGAGAGAC	10	420
C1127205	14	00111001101010101100	CGTGTCTCTA	10	421
C1127205	14	11010100110010101010	TAGATATCTC	10	422
C1127205	14	11001011011001011000	TATCGACAGT	10	423
C1127205	14	11001001100110010011	TATGTGTGCG	10	424
C1127205	14	10110100101110010100	TCGATCGTGA	10	425
C1127205	14	10011110100101001001	TGTACTGATG	10	426
C1127206	14	01100101011001001101	ACAGACATAG	10	427
C1127206	14	01101011001100101001	ACTCGCGCTG	10	428
C1127206	14	01101010110010010110	ACTCTATGAC	10	429
C1127206	14	01010011001111100100	AGCGCGTACA	10	430
C1127206	14	01001110010110110100	ATACAGTCGA	10	431
C1127206	14	00100111110111001000	CACGTAGTAT	10	432
C1127206	14	00110101001001110011	CGAGCACGCG	10	433
C1127206	14	00111001010101010110	CGTGAGAGAC	10	434
C1127206	14	00111001101010101100	CGTGTCTCTA	10	435
C1127206	14	11010010110100101010	TAGCTAGCTC	10	436
C1127206	14	11001011011001011000	TATCGACAGT	10	437
C1127206	14	11001001100110010011	TATGTGTGCG	10	438
C1127206	14	10110100101110010100	TCGATCGTGA	10	439
C1127206	14	10011110100101001001	TGTACTGATG	10	440
C1127207	14	01100101011001001101	ACAGACATAG	10	441
C1127207	14	01101011001100101001	ACTCGCGCTG	10	442
C1127207	14	01101010110010010110	ACTCTATGAC	10	443
C1127207	14	01010011001111100100	AGCGCGTACA	10	444
C1127207	14	01001110010110110100	ATACAGTCGA	10	445
C1127207	14	00100111110111001000	CACGTAGTAT	10	446
C1127207	14	00110101001001110011	CGAGCACGCG	10	447
C1127207	14	00111001010101010110	CGTGAGAGAC	10	448
C1127207	14	00111001101010101100	CGTGTCTCTA	10	449
C1127207	14	11010100110010101010	TAGATATCTC	10	450
C1127207	14	11010011011001011000	TAGCGACAGT	10	451
C1127207	14	11001001100110010011	TATGTGTGCG	10	452
C1127207	14	10110100101110010100	TCGATCGTGA	10	453
C1127207	14	10011110100101001001	TGTACTGATG	10	454
C1127208	14	01100100110011110010	ACATATACGC	10	455
C1127208	14	01101011001100101001	ACTCGCGCTG	10	456

IDT Exhibit 1019.0243

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127208	14	01010011001111100100	AGCGCGTACA	10	457
C1127208	14	01011110100110011000	AGTACTGTGT	10	458
C1127208	14	00100110010111011100	CACAGTAGTA	10	459
C1127208	14	00110101001001110011	CGAGCACGCG	10	460
C1127208	14	00111001010101010110	CGTGAGAGAC	10	461
C1127208	14	00111001101010101100	CGTGTCTCTA	10	462
C1127208	14	00101010111001001011	CTCTACATCG	10	463
C1127208	14	11010010110100101010	TAGCTAGCTC	10	464
C1127208	14	11001011011001011000	TATCGACAGT	10	465
C1127208	14	11001001100110010011	TATGTGTGCG	10	466
C1127208	14	10110100101110010100	TCGATCGTGA	10	467
C1127208	14	10010010011010110110	TGCACTCGAC	10	468
C1127209	14	01101011001100101001	ACTCGCGCTG	10	469
C1127209	14	01010011001111100100	AGCGCGTACA	10	470
C1127209	14	01001101101011011000	ATAGTCTAGT	10	471
C1127209	14	01001010010011100111	ATCATACACG	10	472
C1127209	14	00100111110111001000	CACGTAGTAT	10	473
C1127209	14	00110101001001110011	CGAGCACGCG	10	474
C1127209	14	00111001010101010110	CGTGAGAGAC	10	475
C1127209	14	00111001101010101100	CGTGTCTCTA	10	476
C1127209	14	00101110100101110100	CTACTGACGA	10	477
C1127209	14	11010100110010101010	TAGATATCTC	10	478
C1127209	14	11001111001001001100	TATACGCATA	10	479
C1127209	14	11001001100110010011	TATGTGTGCG	10	480
C1127209	14	10110100101110010100	TCGATCGTGA	10	481
C1127209	14	10010010011010110110	TGCACTCGAC	10	482
C1127210	14	01100101011001001101	ACAGACATAG	10	483
C1127210	14	01101011001100101001	ACTCGCGCTG	10	484
C1127210	14	01010011001111100100	AGCGCGTACA	10	485
C1127210	14	01011100111001011000	AGTATACAGT	10	486
C1127210	14	01001110010110110100	ATACAGTCGA	10	487
C1127210	14	00100111110111001000	CACGTAGTAT	10	488
C1127210	14	00110101001001110011	CGAGCACGCG	10	489
C1127210	14	00111001010101010110	CGTGAGAGAC	10	490
C1127210	14	00111001101010101100	CGTGTCTCTA	10	491
C1127210	14	00101010010010101111	CTCATCTACG	10	492
C1127210	14	11010100110010101010	TAGATATCTC	10	493
C1127210	14	11001001100110010011	TATGTGTGCG	10	494
C1127210	14	10110010101101011000	TCGCTCGAGT	10	495
C1127210	14	10101110100100100110	TCTACTGCAC	10	496
C1127211	14	01100101011001001101	ACAGACATAG	10	497
C1127211	14	01101011001100101001	ACTCGCGCTG	10	498
C1127211	14	01010011001111100100	AGCGCGTACA	10	499
C1127211	14	01011100111001011000	AGTATACAGT	10	500
C1127211	14	01001110010110110100	ATACAGTCGA	10	501

IDT Exhibit 1019.0244

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127211	14	00100111110111001000	CACGTAGTAT	10	502
C1127211	14	00110101001001110011	CGAGCACGCG	10	503
C1127211	14	00111001010101010110	CGTGAGAGAC	10	504
C1127211	14	00111001101010101100	CGTGTCTCTA	10	505
C1127211	14	00101010010010101111	CTCATCTACG	10	506
C1127211	14	11010100110010101010	TAGATATCTC	10	507
C1127211	14	11001001100110010011	TATGTGTGCG	10	508
C1127211	14	10110010101101011000	TCGCTCGAGT	10	509
C1127211	14	10011110100101001001	TGTACTGATG	10	510
C1127212	14	01100101011001001101	ACAGACATAG	10	511
C1127212	14	01101011001100101001	ACTCGCGCTG	10	512
C1127212	14	01101010110010010110	ACTCTATGAC	10	513
C1127212	14	01010011001111100100	AGCGCGTACA	10	514
C1127212	14	01011100111001011000	AGTATACAGT	10	515
C1127212	14	01001110010110110100	ATACAGTCGA	10	516
C1127212	14	00100111110111001000	CACGTAGTAT	10	517
C1127212	14	00110101001001110011	CGAGCACGCG	10	518
C1127212	14	00111001010101010110	CGTGAGAGAC	10	519
C1127212	14	00111001101010101100	CGTGTCTCTA	10	520
C1127212	14	11010100110010101010	TAGATATCTC	10	521
C1127212	14	11001001100110010011	TATGTGTGCG	10	522
C1127212	14	10110010101101011000	TCGCTCGAGT	10	523
C1127212	14	10011110100101001001	TGTACTGATG	10	524
C1127213	14	01100111011010010010	ACACGACTGC	10	525
C1127213	14	01011100111001011000	AGTATACAGT	10	526
C1127213	14	01001110101001001011	ATACTCATCG	10	527
C1127213	14	01001011010011101100	ATCGATACTA	10	528
C1127213	14	00110101001001110011	CGAGCACGCG	10	529
C1127213	14	00111001010101010110	CGTGAGAGAC	10	530
C1127213	14	00111001101010101100	CGTGTCTCTA	10	531
C1127213	14	00101110010110011100	CTACAGTGTA	10	532
C1127213	14	00101010111100100110	CTCTACGCAC	10	533
C1127213	14	11010010101010010110	TAGCTCTGAC	10	534
C1127213	14	10100110010011010101	TCACATAGAG	10	535
C1127213	14	10101001100110010011	TCTGTGTGCG	10	536
C1127213	14	10010100110110110010	TGATAGTCGC	10	537
C1127213	14	10011011001101001001	TGTCGCGATG	10	538
C1127214	14	01100110101010011010	ACACTCTGTC	10	539
C1127214	14	01011100111001011000	AGTATACAGT	10	540
C1127214	14	01001101001110100101	ATAGCGTCAG	10	541
C1127214	14	01001011010011101100	ATCGATACTA	10	542
C1127214	14	00110101001001110011	CGAGCACGCG	10	543
C1127214	14	00111001010101010110	CGTGAGAGAC	10	544
C1127214	14	00111001101010101100	CGTGTCTCTA	10	545
C1127214	14	00101111010110011000	CTACGAGTGT	10	546

IDT Exhibit 1019.0245

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127214	14	00101010111100100110	CTCTACGCAC	10	547
C1127214	14	11101001011001001010	TACTGACATC	10	548
C1127214	14	10100110010011010101	TCACATAGAG	10	549
C1127214	14	10101001100110010011	TCTGTGTGCG	10	550
C1127214	14	10010100110110110010	TGATAGTCGC	10	551
C1127214	14	10011011001101001001	TGTCGCGATG	10	552
C1127215	14	01100100111010111000	ACATACTCGT	10	553
C1127215	14	01010010101111010100	AGCTCGTAGA	10	554
C1127215	14	01011100100110010011	AGTATGTGCG	10	555
C1127215	14	01001010100101111010	ATCTGACGTC	10	556
C1127215	14	00110101001001110011	CGAGCACGCG	10	557
C1127215	14	00111001010101010110	CGTGAGAGAC	10	558
C1127215	14	00111001101010101100	CGTGTCTCTA	10	559
C1127215	14	00101111010110011000	CTACGAGTGT	10	560
C1127215	14	00101010111001100101	CTCTACACAG	10	561
C1127215	14	11100100110101001100	TACATAGATA	10	562
C1127215	14	11010011100100100110	TAGCGTGCAC	10	563
C1127215	14	10100110100110101001	TCACTGTCTG	10	564
C1127215	14	10111010010011101000	TCGTCATACT	10	565
C1127215	14	10011011001101001001	TGTCGCGATG	10	566
C1127216	14	01100100111010110010	ACATACTCGC	10	567
C1127216	14	01010010101111010100	AGCTCGTAGA	10	568
C1127216	14	01011100100110010011	AGTATGTGCG	10	569
C1127216	14	01001010100101111010	ATCTGACGTC	10	570
C1127216	14	00110101001001110011	CGAGCACGCG	10	571
C1127216	14	00111001010101010110	CGTGAGAGAC	10	572
C1127216	14	00111001101010101100	CGTGTCTCTA	10	573
C1127216	14	00101111010110011000	CTACGAGTGT	10	574
C1127216	14	00101010111001100101	CTCTACACAG	10	575
C1127216	14	11100100110101001100	TACATAGATA	10	576
C1127216	14	11010011100100100110	TAGCGTGCAC	10	577
C1127216	14	10100110100110101001	TCACTGTCTG	10	578
C1127216	14	10111010010011101000	TCGTCATACT	10	579
C1127216	14	10011011001101001001	TGTCGCGATG	10	580
C1127217	14	01100110011001101010	ACACACACTC	10	581
C1127217	14	01100100101010011101	ACATCTGTAG	10	582
C1127217	14	01110011001110011000	ACGCGCGTGT	10	583
C1127217	14	01010011010011001110	AGCGATATAC	10	584
C1127217	14	01011101100101010100	AGTAGTGAGA	10	585
C1127217	14	00110101010100110011	CGAGAGCGCG	10	586
C1127217	14	00111010010011111000	CGTCATACGT	10	587
C1127217	14	00111001101010101100	CGTGTCTCTA	10	588
C1127217	14	00101110110110101000	CTACTAGTCT	10	589
C1127217	14	00101011001101100101	CTCGCGACAG	10	590
C1127217	14	11010110101101001000	TAGACTCGAT	10	591

IDT Exhibit 1019.0246

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127217	14	11001001010101011010	TATGAGAGTC	10	592
C1127217	14	10010010100111010011	TGCTGTAGCG	10	593
C1127217	14	10011111001010010010	TGTACGCTGC	10	594
C1127218	14	01100100101010011101	ACATCTGTAG	10	595
C1127218	14	01110011001110011000	ACGCGCGTGT	10	596
C1127218	14	01010011010011001110	AGCGATATAC	10	597
C1127218	14	01011101100101010100	AGTAGTGAGA	10	598
C1127218	14	00110101010100110011	CGAGAGCGCG	10	599
C1127218	14	00111010010011111000	CGTCATACGT	10	600
C1127218	14	00111001101010101100	CGTGTCTCTA	10	601
C1127218	14	00101110110110101000	CTACTAGTCT	10	602
C1127218	14	00101011001101100101	CTCGCGACAG	10	603
C1127218	14	11010110101101001000	TAGACTCGAT	10	604
C1127218	14	11001100111001100100	TATATACACA	10	605
C1127218	14	10100110010101011100	TCACAGAGTA	10	606
C1127218	14	10010010100111010011	TGCTGTAGCG	10	607
C1127218	14	10011111001010010010	TGTACGCTGC	10	608
C1127219	14	01110011001110011000	ACGCGCGTGT	10	609
C1127219	14	01010101011001001110	AGAGACATAC	10	610
C1127219	14	01001101110110101000	ATAGTAGTCT	10	611
C1127219	14	01001010100101110101	ATCTGACGAG	10	612
C1127219	14	00100111001011001011	CACGCTATCG	10	613
C1127219	14	00110101010100110011	CGAGAGCGCG	10	614
C1127219	14	00111010010011111000	CGTCATACGT	10	615
C1127219	14	00111001101010101100	CGTGTCTCTA	10	616
C1127219	14	00101010011110100110	CTCACGTCAC	10	617
C1127219	14	11010010101010010110	TAGCTCTGAC	10	618
C1127219	14	11001110010010011001	TATACATGTG	10	619
C1127219	14	10100110100101011100	TCACTGAGTA	10	620
C1127219	14	10101100111001100100	TCTATACACA	10	621
C1127219	14	10010011110101001001	TGCGTAGATG	10	622
C1127220	14	01100110010011001101	ACACATATAG	10	623
C1127220	14	01110011001110011000	ACGCGCGTGT	10	624
C1127220	14	01010111001001101010	AGACGCACTC	10	625
C1127220	14	01001100111110100100	ATATACGTCA	10	626
C1127220	14	01001001101011010011	ATGTCTAGCG	10	627
C1127220	14	00110101010100110011	CGAGAGCGCG	10	628
C1127220	14	00110010101101100101	CGCTCGACAG	10	629
C1127220	14	00111001101010101100	CGTGTCTCTA	10	630
C1127220	14	00101011110101001010	CTCGTAGATC	10	631
C1127220	14	11101001010100101100	TACTGAGCTA	10	632
C1127220	14	11010011010010010110	TAGCGATGAC	10	633
C1127220	14	10101100100110011001	TCTATGTGTG	10	634
C1127220	14	10101010011010110010	TCTCACTCGC	10	635
C1127220	14	10010101111001011000	TGAGTACAGT	10	636

IDT Exhibit 1019.0247

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127221	14	01100110010011001101	ACACATATAG	10	637
C1127221	14	01110011001110011000	ACGCGCGTGT	10	638
C1127221	14	01010111001001101010	AGACGCACTC	10	639
C1127221	14	01001100111110100100	ATATACGTCA	10	640
C1127221	14	01001001101011010011	ATGTCTAGCG	10	641
C1127221	14	00110101010100110011	CGAGAGCGCG	10	642
C1127221	14	00110010101101100101	CGCTCGACAG	10	643
C1127221	14	00111001101010101100	CGTGTCTCTA	10	644
C1127221	14	00101011110101001010	CTCGTAGATC	10	645
C1127221	14	11101001010100101100	TACTGAGCTA	10	646
C1127221	14	11010011010010010110	TAGCGATGAC	10	647
C1127221	14	10101100100110011001	TCTATGTGTG	10	648
C1127221	14	10101010011010110010	TCTCACTCGC	10	649
C1127221	14	10010101111001010100	TGAGTACAGA	10	650
C1127222	14	01100100110011100101	ACATATACAG	10	651
C1127222	14	01110011001110011000	ACGCGCGTGT	10	652
C1127222	14	01010111001001101010	AGACGCACTC	10	653
C1127222	14	01001001011011010011	ATGACTAGCG	10	654
C1127222	14	00100110010111011100	CACAGTAGTA	10	655
C1127222	14	00110101010100110011	CGAGAGCGCG	10	656
C1127222	14	00110010111010010110	CGCTACTGAC	10	657
C1127222	14	00111001101010101100	CGTGTCTCTA	10	658
C1127222	14	00101011001101100101	CTCGCGACAG	10	659
C1127222	14	11100100101101010010	TACATCGAGC	10	660
C1127222	14	11011110101110000000	TAGTACTCGT	10	661
C1127222	14	10101110011010100100	TCTACACTCA	10	662
C1127222	14	10101100100110011001	TCTATGTGTG	10	663
C1127222	14	10011010010101001110	TGTCAGATAC	10	664
C1127223	14	01100110011001101010	ACACACACTC	10	665
C1127223	14	01100100101010011101	ACATCTGTAG	10	666
C1127223	14	01110011001110011000	ACGCGCGTGT	10	667
C1127223	14	01010011010011001110	AGCGATATAC	10	668
C1127223	14	01011101100101010100	AGTAGTGAGA	10	669
C1127223	14	00110101010100110011	CGAGAGCGCG	10	670
C1127223	14	00111001101010101100	CGTGTCTCTA	10	671
C1127223	14	00101110110110101000	CTACTAGTCT	10	672
C1127223	14	00101011001101100101	CTCGCGACAG	10	673
C1127223	14	11010110101101001000	TAGACTCGAT	10	674
C1127223	14	11001010010010100111	TATCATCACG	10	675
C1127223	14	11001001010101011010	TATGAGAGTC	10	676
C1127223	14	10010010100111010011	TGCTGTAGCG	10	677
C1127223	14	10011111001010010010	TGTACGCTGC	10	678
C1127224	14	01100110101010100101	ACACTCTCAG	10	679
C1127224	14	01110011001110011000	ACGCGCGTGT	10	680
C1127224	14	01111101011001001000	ACGTAGACAT	10	681

IDT Exhibit 1019.0248

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127224	14	01010110010111001100	AGACAGTATA	10	682
C1127224	14	01011100100110010011	AGTATGTGCG	10	683
C1127224	14	00110101001100100111	CGAGCGCACG	10	684
C1127224	14	00111001101010101100	CGTGTCTCTA	10	685
C1127224	14	00101111010010101010	CTACGATCTC	10	686
C1127224	14	00101010111101010100	CTCTACGAGA	10	687
C1127224	14	11011010011100100100	TAGTCACGCA	10	688
C1127224	14	11001101001001011001	TATAGCAGTG	10	689
C1127224	14	10111010100101001010	TCGTCTGATC	10	690
C1127224	14	10101001010011001101	TCTGATATAG	10	691
C1127224	14	10010010011011111000	TGCACTACGT	10	692
C1127225	14	01100101101100100110	ACAGTCGCAC	10	693
C1127225	14	01110011001110011000	ACGCGCGTGT	10	694
C1127225	14	01010110100110010101	AGACTGTGAG	10	695
C1127225	14	01011100111001001001	AGTATACATG	10	696
C1127225	14	00100110111011010010	CACTACTAGC	10	697
C1127225	14	00110101001001110011	CGAGCACGCG	10	698
C1127225	14	00111001101010101100	CGTGTCTCTA	10	699
C1127225	14	00101111001101001100	CTACGCGATA	10	700
C1127225	14	00101010010110011011	CTCAGTGTCG	10	701
C1127225	14	11001011001010100101	TATCGCTCAG	10	702
C1127225	14	11001001110110010100	TATGTAGTGA	10	703
C1127225	14	10110010010011110100	TCGCATACGA	10	704
C1127225	14	10010101010011001110	TGAGATATAC	10	705
C1127225	14	10011011010101101000	TGTCGAGACT	10	706
C1127226	14	01100100111011001001	ACATACTATG	10	707
C1127226	14	01110011001110011000	ACGCGCGTGT	10	708
C1127226	14	01011101010011011000	AGTAGATAGT	10	709
C1127226	14	01001001100101110110	ATGTGACGAC	10	710
C1127226	14	00100101010010111110	CAGATCGTAC	10	711
C1127226	14	00110101001001110011	CGAGCACGCG	10	712
C1127226	14	00111001101010101100	CGTGTCTCTA	10	713
C1127226	14	00101111001101001100	CTACGCGATA	10	714
C1127226	14	00101010010110011011	CTCAGTGTCG	10	715
C1127226	14	11010110011010010100	TAGACACTGA	10	716
C1127226	14	11001010110100111000	TATCTAGCGT	10	717
C1127226	14	10111100100101010100	TCGTATGAGA	10	718
C1127226	14	10101011001010100110	TCTCGCTCAC	10	719
C1127226	14	10010010110011101010	TGCTATACTC	10	720
C1127227	14	01100101101100100110	ACAGTCGCAC	10	721
C1127227	14	01110011001110011000	ACGCGCGTGT	10	722
C1127227	14	01011100110101100100	AGTATAGACA	10	723
C1127227	14	01001001011011100101	ATGACTACAG	10	724
C1127227	14	00100111110101001001	CACGTAGATG	10	725
C1127227	14	00110101010100110101	CGAGAGCGAG	10	726

IDT Exhibit 1019.0249

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127227	14	00111010010011111000	CGTCATACGT	10	727
C1127227	14	00111001101010101100	CGTGTCTCTA	10	728
C1127227	14	00101010011100101011	CTCACGCTCG	10	729
C1127227	14	11110100101001101000	TACGATCACT	10	730
C1127227	14	11010010010010011101	TAGCATGTAG	10	731
C1127227	14	11001110010100110010	TATACAGCGC	10	732
C1127227	14	10110101010011001010	TCGAGATATC	10	733
C1127227	14	10101001100110010011	TCTGTGTGCG	10	734
C1127228	14	01100101101100100110	ACAGTCGCAC	10	735
C1127228	14	01110011001110011000	ACGCGCGTGT	10	736
C1127228	14	01010110100101011010	AGACTGAGTC	10	737
C1127228	14	01001010111001010101	ATCTACAGAG	10	738
C1127228	14	00100111110101001001	CACGTAGATG	10	739
C1127228	14	00110101010100110101	CGAGAGCGAG	10	740
C1127228	14	00111010010011111000	CGTCATACGT	10	741
C1127228	14	00111001101010101100	CGTGTCTCTA	10	742
C1127228	14	00101010011100101011	CTCACGCTCG	10	743
C1127228	14	11110100101001101000	TACGATCACT	10	744
C1127228	14	11010010010010011101	TAGCATGTAG	10	745
C1127228	14	11001111001001001100	TATACGCATA	10	746
C1127228	14	10110101010011001010	TCGAGATATC	10	747
C1127228	14	10101001100110010011	TCTGTGTGCG	10	748
C1127229	14	01100110110010101001	ACACTATCTG	10	749
C1127229	14	01110011001110011000	ACGCGCGTGT	10	750
C1127229	14	01010011010011001110	AGCGATATAC	10	751
C1127229	14	01011010101101110000	AGTCTCGACG	10	752
C1127229	14	01001110010101011010	ATACAGAGTC	10	753
C1127229	14	00100111100101110010	CACGTGACGC	10	754
C1127229	14	00110101010100110101	CGAGAGCGAG	10	755
C1127229	14	00111001101010101100	CGTGTCTCTA	10	756
C1127229	14	00101010110111010100	CTCTAGTAGA	10	757
C1127229	14	11011010010111001000	TAGTCAGTAT	10	758
C1127229	14	11001100111001100100	TATATACACA	10	759
C1127229	14	10101011001010100110	TCTCGCTCAC	10	760
C1127229	14	10101001100110010011	TCTGTGTGCG	10	761
C1127229	14	10010101001011111000	TGAGCTACGT	10	762
C1127230	14	01110011001110010100	ACGCGCGTGA	10	763
C1127230	14	01101001110101010010	ACTGTAGAGC	10	764
C1127230	14	01010101011001001110	AGAGACATAC	10	765
C1127230	14	01001111100100101010	ATACGTGCTC	10	766
C1127230	14	00100110111001101001	CACTACACTG	10	767
C1127230	14	00110101010100110011	CGAGAGCGCG	10	768
C1127230	14	00111010010011111000	CGTCATACGT	10	769
C1127230	14	00111001101010101100	CGTGTCTCTA	10	770
C1127230	14	00101010010110011101	CTCAGTGTAG	10	771

IDT Exhibit 1019.0250

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127230	14	11010100110110011000	TAGATAGTGT	10	772
C1127230	14	11001011011100100100	TATCGACGCA	10	773
C1127230	14	10101111010011001000	TCTACGATAT	10	774
C1127230	14	10101100101010010011	TCTATCTGCG	10	775
C1127230	14	10010010011010110110	TGCACTCGAC	10	776
C1127231	14	01110011001110010100	ACGCGCGTGA	10	777
C1127231	14	01010111001001101010	AGACGCACTC	10	778
C1127231	14	01011101010111001000	AGTAGAGTAT	10	779
C1127231	14	01001011100101100101	ATCGTGACAG	10	780
C1127231	14	01001001011011010011	ATGACTAGCG	10	781
C1127231	14	00100110011010011101	CACACTGTAG	10	782
C1127231	14	00110101010100110011	CGAGAGCGCG	10	783
C1127231	14	00111010010011111000	CGTCATACGT	10	784
C1127231	14	00111001101010101100	CGTGTCTCTA	10	785
C1127231	14	11011010011100100100	TAGTCACGCA	10	786
C1127231	14	11001010110101001010	TATCTAGATC	10	787
C1127231	14	10100111010011100100	TCACGATACA	10	788
C1127231	14	10101100100110011001	TCTATGTGTG	10	789
C1127231	14	10010010010110110110	TGCAGTCGAC	10	790
C1127232	14	01110011001110010100	ACGCGCGTGA	10	791
C1127232	14	01101001010111101000	ACTGAGTACT	10	792
C1127232	14	01010100101111001010	AGATCGTATC	10	793
C1127232	14	01001110101010100101	ATACTCTCAG	10	794
C1127232	14	01001001111001010011	ATGTACAGCG	10	795
C1127232	14	00110101010100110011	CGAGAGCGCG	10	796
C1127232	14	00110010010111001101	CGCAGTATAG	10	797
C1127232	14	00111001101010101100	CGTGTCTCTA	10	798
C1127232	14	00101111011001001100	CTACGACATA	10	799
C1127232	14	11010101011010011000	TAGAGACTGT	10	800
C1127232	14	11010011100100100110	TAGCGTGCAC	10	801
C1127232	14	10111010110101001000	TCGTCTAGAT	10	802
C1127232	14	10101100100110011001	TCTATGTGTG	10	803
C1127232	14	10101010011010110010	TCTCACTCGC	10	804
C1127233	14	01100101101100100110	ACAGTCGCAC	10	805
C1127233	14	01101001011001011001	ACTGACAGTG	10	806
C1127233	14	01010100110101101100	AGATAGACTA	10	807
C1127233	14	01010011001011100101	AGCGCTACAG	10	808
C1127233	14	01001010110010101011	ATCTATCTCG	10	809
C1127233	14	00110101010100110011	CGAGAGCGCG	10	810
C1127233	14	00111010010011111000	CGTCATACGT	10	811
C1127233	14	00111001101010101100	CGTGTCTCTA	10	812
C1127233	14	00101111100101100100	CTACGTGACA	10	813
C1127233	14	11011110100100101000	TAGTACTGCT	10	814
C1127233	14	11001010101001010110	TATCTCAGAC	10	815
C1127233	14	10100101010111010100	TCAGAGTAGA	10	816

IDT Exhibit 1019.0251

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127233	14	10110010100110011001	TCGCTGTGTG	10	817
C1127233	14	10010011011101001010	TGCGACGATC	10	818
C1127234	14	01100110011010011001	ACACACTGTG	10	819
C1127234	14	01110010110101001100	ACGCTAGATA	10	820
C1127234	14	01101001100100111010	ACTGTGCGTC	10	821
C1127234	14	01001111001001110100	ATACGCACGA	10	822
C1127234	14	00110101010100110011	CGAGAGCGCG	10	823
C1127234	14	00110010011110100110	CGCACGTCAC	10	824
C1127234	14	00111010010011111000	CGTCATACGT	10	825
C1127234	14	00111001101010101100	CGTGTCTCTA	10	826
C1127234	14	00101110110110101000	CTACTAGTCT	10	827
C1127234	14	11100100101101010010	TACATCGAGC	10	828
C1127234	14	11011001110011100000	TAGTGTATAC	10	829
C1127234	14	10101100100101100101	TCTATGACAG	10	830
C1127234	14	10010100111001101010	TGATACACTC	10	831
C1127234	14	10011011010110010100	TGTCGAGTGA	10	832
C1127235	14	01110100101010010101	ACGATCTGAG	10	833
C1127235	14	01101011001100101001	ACTCGCGCTG	10	834
C1127235	14	01011100110101100100	AGTATAGACA	10	835
C1127235	14	01001110101001001011	ATACTCATCG	10	836
C1127235	14	01001001100101111010	ATGTGACGTC	10	837
C1127235	14	00100111110111001000	CACGTAGTAT	10	838
C1127235	14	00110101010100110011	CGAGAGCGCG	10	839
C1127235	14	00110010011110100110	CGCACGTCAC	10	840
C1127235	14	00111010010011111000	CGTCATACGT	10	841
C1127235	14	00111001101010101100	CGTGTCTCTA	10	842
C1127235	14	11010100110010101010	TAGATATCTC	10	843
C1127235	14	11010010011101011000	TAGCACGAGT	10	844
C1127235	14	11001101100110010100	TATAGTGTGA	10	845
C1127235	14	10101010100101001110	TCTCTGATAC	10	846
C1127236	14	01100100101010111001	ACATCTCGTG	10	847
C1127236	14	01110010100101100101	ACGCTGACAG	10	848
C1127236	14	01010111001001101010	AGACGCACTC	10	849
C1127236	14	01011110110010011000	AGTACTATGT	10	850
C1127236	14	01001011010101011100	ATCGAGAGTA	10	851
C1127236	14	00100101001110011110	CAGCGTGTAC	10	852
C1127236	14	00110101010100110011	CGAGAGCGCG	10	853
C1127236	14	00111001101010101100	CGTGTCTCTA	10	854
C1127236	14	00101010011001110110	CTCACACGAC	10	855
C1127236	14	11100101010011010010	TACAGATAGC	10	856
C1127236	14	11011010011100100100	TAGTCACGCA	10	857
C1127236	14	10101110100100101010	TCTACTGCTC	10	858
C1127236	14	10010011001110111000	TGCGCGTCGT	10	859
C1127236	14	10011100101001010101	TGTATCAGAG	10	860
C1127237	14	01101010110010100110	ACTCTATCAC	10	861

IDT Exhibit 1019.0252

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127237	14	01101001010011011001	ACTGATAGTG	10	862
C1127237	14	01010111001001101010	AGACGCACTC	10	863
C1127237	14	01011101100101010100	AGTAGTGAGA	10	864
C1127237	14	00100101001110011110	CAGCGTGTAC	10	865
C1127237	14	00110101010100110011	CGAGAGCGCG	10	866
C1127237	14	00111001101010101100	CGTGTCTCTA	10	867
C1127237	14	00101010101101010101	CTCTCGAGAG	10	868
C1127237	14	11100101011101100000	TACAGACGAC	10	869
C1127237	14	11011101010010101000	TAGTAGATCT	10	870
C1127237	14	11001010011010010011	TATCACTGCG	10	871
C1127237	14	10101110011001001100	TCTACACATA	10	872
C1127237	14	10010100100111001101	TGATGTATAG	10	873
C1127237	14	10010011001110111000	TGCGCGTCGT	10	874
C1127238	14	01101010110010100110	ACTCTATCAC	10	875
C1127238	14	01101001010011011001	ACTGATAGTG	10	876
C1127238	14	01010111001001101010	AGACGCACTC	10	877
C1127238	14	01011101100101010100	AGTAGTGAGA	10	878
C1127238	14	00100101001110011110	CAGCGTGTAC	10	879
C1127238	14	00110101010100110011	CGAGAGCGCG	10	880
C1127238	14	00111001101010101100	CGTGTCTCTA	10	881
C1127238	14	00101010101101010101	CTCTCGAGAG	10	882
C1127238	14	11100110100100101001	TACACTGCTG	10	883
C1127238	14	11100101011101100000	TACAGACGAC	10	884
C1127238	14	11011101010010101000	TAGTAGATCT	10	885
C1127238	14	10101110011001001100	TCTACACATA	10	886
C1127238	14	10010100100111001101	TGATGTATAG	10	887
C1127238	14	10010011001110111000	TGCGCGTCGT	10	888
C1127239	14	01110010010010101110	ACGCATCTAC	10	889
C1127239	14	01101011001100101001	ACTCGCGCTG	10	890
C1127239	14	01010110100110011010	AGACTGTGTC	10	891
C1127239	14	01001100111100100110	ATATACGCAC	10	892
C1127239	14	00100111110110010100	CACGTAGTGA	10	893
C1127239	14	00110101010100110011	CGAGAGCGCG	10	894
C1127239	14	00111011001001010110	CGTCGCAGAC	10	895
C1127239	14	00111001101010101100	CGTGTCTCTA	10	896
C1127239	14	00101010111001001011	CTCTACATCG	10	897
C1127239	14	11100101010011010010	TACAGATAGC	10	898
C1127239	14	11011010100101100100	TAGTCTGACA	10	899
C1127239	14	10101100101101011000	TCTATCGAGT	10	900
C1127239	14	10010100110011101001	TGATATACTG	10	901
C1127239	14	10010011001110111000	TGCGCGTCGT	10	902
C1127240	14	01101011100110010100	ACTCGTGTGA	10	903
C1127240	14	01101001011001111000	ACTGACACGT	10	904
C1127240	14	01010010011011101010	AGCACTACTC	10	905
C1127240	14	01001110010101001110	ATACAGATAC	10	906

IDT Exhibit 1019.0253

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127240	14	01001100100110111010	ATATGTCGTC	10	907
C1127240	14	00110101010100110011	CGAGAGCGCG	10	908
C1127240	14	00111011001001010110	CGTCGCAGAC	10	909
C1127240	14	00111001101010101100	CGTGTCTCTA	10	910
C1127240	14	00101010111111001000	CTCTACGTAT	10	911
C1127240	14	11011010011100100100	TAGTCACGCA	10	912
C1127240	14	10100110011011010100	TCACACTAGA	10	913
C1127240	14	10101100101010010011	TCTATCTGCG	10	914
C1127240	14	10010100100111001101	TGATGTATAG	10	915
C1127240	14	10010011001110111000	TGCGCGTCGT	10	916
C1127241	14	01101011100110010100	ACTCGTGTGA	10	917
C1127241	14	01101001001010100111	ACTGCTCACG	10	918
C1127241	14	01011100111001011000	AGTATACAGT	10	919
C1127241	14	01001110010111100100	ATACAGTACA	10	920
C1127241	14	01001101101101001010	ATAGTCGATC	10	921
C1127241	14	00100111010011101010	CACGATACTC	10	922
C1127241	14	00110101010100110011	CGAGAGCGCG	10	923
C1127241	14	00111011001001010110	CGTCGCAGAC	10	924
C1127241	14	00111001101010101100	CGTGTCTCTA	10	925
C1127241	14	00101010111111001000	CTCTACGTAT	10	926
C1127241	14	11001010110100110010	TATCTAGCGC	10	927
C1127241	14	11001001010101001101	TATGAGATAG	10	928
C1127241	14	10100110011010010101	TCACACTGAG	10	929
C1127241	14	10010011001110111000	TGCGCGTCGT	10	930
C1127242	14	01101011100110010100	ACTCGTGTGA	10	931
C1127242	14	01101001001010100111	ACTGCTCACG	10	932
C1127242	14	01011100111001011000	AGTATACAGT	10	933
C1127242	14	01001101101101001010	ATAGTCGATC	10	934
C1127242	14	00100111010011101010	CACGATACTC	10	935
C1127242	14	00110101010100110011	CGAGAGCGCG	10	936
C1127242	14	00111011001001010110	CGTCGCAGAC	10	937
C1127242	14	00111001101010101100	CGTGTCTCTA	10	938
C1127242	14	00101010111111001000	CTCTACGTAT	10	939
C1127242	14	11001010110100110010	TATCTAGCGC	10	940
C1127242	14	11001001010101001101	TATGAGATAG	10	941
C1127242	14	10100110011011010100	TCACACTAGA	10	942
C1127242	14	10100101110010011001	TCAGTATGTG	10	943
C1127242	14	10010011001110111000	TGCGCGTCGT	10	944
C1127243	14	01100100111010100110	ACATACTCAC	10	945
C1127243	14	01110010011101001100	ACGCACGATA	10	946
C1127243	14	01101001100110010101	ACTGTGTGAG	10	947
C1127243	14	01011100111001011000	AGTATACAGT	10	948
C1127243	14	01001111110010010010	ATACGTATGC	10	949
C1127243	14	00110101010100110011	CGAGAGCGCG	10	950
C1127243	14	00111011001001010110	CGTCGCAGAC	10	951

IDT Exhibit 1019.0254

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127243	14	00111001101010101100	CGTGTCTCTA	10	952
C1127243	14	00101010111001001011	CTCTACATCG	10	953
C1127243	14	11100110100100101001	TACACTGCTG	10	954
C1127243	14	11001001010101101010	TATGAGACTC	10	955
C1127243	14	10100101001111100100	TCAGCGTACA	10	956
C1127243	14	10101011001100111000	TCTCGCGCGT	10	957
C1127243	14	10010110110111001000	TGACTAGTAT	10	958
C1127244	14	01110010110101101000	ACGCTAGACT	10	959
C1127244	14	01011011100101010100	AGTCGTGAGA	10	960
C1127244	14	01001100101001001111	ATATCATACG	10	961
C1127244	14	00100100110111011001	CATAGTAGTG	10	962
C1127244	14	00110101010101001110	CGAGAGATAC	10	963
C1127244	14	00111001101010101100	CGTGTCTCTA	10	964
C1127244	14	00101011001110011010	CTCGCGTGTC	10	965
C1127244	14	00101010111100100101	CTCTACGCAG	10	966
C1127244	14	11100110011001001100	TACACACATA	10	967
C1127244	14	11101001100111001000	TACTGTGTAT	10	968
C1127244	14	11010010101010110010	TAGCTCTCGC	10	969
C1127244	14	10101101010010110100	TCTAGATCGA	10	970
C1127244	14	10010110010100110011	TGACAGCGCG	10	971
C1127244	14	10010101111001011000	TGAGTACAGT	10	972
C1127245	14	01110101100100111000	ACGAGTGCGT	10	973
C1127245	14	01110010101101100100	ACGCTCGACA	10	974
C1127245	14	01010111001001101010	AGACGCACTC	10	975
C1127245	14	01010010011010011101	AGCACTGTAG	10	976
C1127245	14	01001100101100110101	ATATCGCGAG	10	977
C1127245	14	01001010010101100111	ATCAGACACG	10	978
C1127245	14	00100100110111011001	CATAGTAGTG	10	979
C1127245	14	00110101010101001110	CGAGAGATAC	10	980
C1127245	14	00111001101010101100	CGTGTCTCTA	10	981
C1127245	14	00101011001110011010	CTCGCGTGTC	10	982
C1127245	14	11101001010100101100	TACTGAGCTA	10	983
C1127245	14	11011100101001010010	TAGTATCAGC	10	984
C1127245	14	10101010110010010011	TCTCTATGCG	10	985
C1127245	14	10010100111110101000	TGATACGTCT	10	986
C1127246	14	01100110100111001100	ACACTGTATA	10	987
C1127246	14	01100101001101111000	ACAGCGACGT	10	988
C1127246	14	01100100111010100110	ACATACTCAC	10	989
C1127246	14	01001011001100101101	ATCGCGCTAG	10	990
C1127246	14	01001001110011011010	ATGTATAGTC	10	991
C1127246	14	00110110010010111001	CGACATCGTG	10	992
C1127246	14	00110101010101001110	CGAGAGATAC	10	993
C1127246	14	00111001101010101100	CGTGTCTCTA	10	994
C1127246	14	00101010111101010100	CTCTACGAGA	10	995
C1127246	14	11101101010010101000	TACTAGATCT	10	996

IDT Exhibit 1019.0255

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127246	14	11010100100110010101	TAGATGTGAG	10	997
C1127246	14	10101001101001010011	TCTGTCAGCG	10	998
C1127246	14	10010011001110011010	TGCGCGTGTC	10	999
C1127246	14	10011011111001001000	TGTCGTACAT	10	1000
C1127247	14	01100110100111001100	ACACTGTATA	10	1001
C1127247	14	01010110011010111000	AGACACTCGT	10	1002
C1127247	14	01001011001100101101	ATCGCGCTAG	10	1003
C1127247	14	01001001110011011010	ATGTATAGTC	10	1004
C1127247	14	00110101010101001110	CGAGAGATAC	10	1005
C1127247	14	00111010100101010011	CGTCTGAGCG	10	1006
C1127247	14	00111001101010101100	CGTGTCTCTA	10	1007
C1127247	14	00101111101101100000	CTACGTCGAC	10	1008
C1127247	14	11101101001010100100	TACTAGCTCA	10	1009
C1127247	14	11010100100110010101	TAGATGTGAG	10	1010
C1127247	14	10100101101001110010	TCAGTCACGC	10	1011
C1127247	14	10101110010101011000	TCTACAGAGT	10	1012
C1127247	14	10010011001110011010	TGCGCGTGTC	10	1013
C1127247	14	10011011111001001000	TGTCGTACAT	10	1014
C1127248	14	01101010101101100100	ACTCTCGACA	10	1015
C1127248	14	01010100111001101001	AGATACACTG	10	1016
C1127248	14	01001111110010101000	ATACGTATCT	10	1017
C1127248	14	01001001100101111010	ATGTGACGTC	10	1018
C1127248	14	00100111011010010101	CACGACTGAG	10	1019
C1127248	14	00110101010101001110	CGAGAGATAC	10	1020
C1127248	14	00110010011110111000	CGCACGTCGT	10	1021
C1127248	14	00111001101010101100	CGTGTCTCTA	10	1022
C1127248	14	00101001110011010011	CTGTATAGCG	10	1023
C1127248	14	11010100100110011100	TAGATGTGTA	10	1024
C1127248	14	11001010010110100110	TATCAGTCAC	10	1025
C1127248	14	10110110111010100000	TCGACTACTC	10	1026
C1127248	14	10111101001001010100	TCGTAGCAGA	10	1027
C1127248	14	10010101001100110011	TGAGCGCGCG	10	1028
C1127249	14	01100101001011100101	ACAGCTACAG	10	1029
C1127249	14	01010010101010011011	AGCTCTGTCG	10	1030
C1127249	14	01001111100100101010	ATACGTGCTC	10	1031
C1127249	14	01001001110011111000	ATGTATACGT	10	1032
C1127249	14	00100111010110011001	CACGAGTGTG	10	1033
C1127249	14	00110101010101001110	CGAGAGATAC	10	1034
C1127249	14	00111001101010101100	CGTGTCTCTA	10	1035
C1127249	14	00101010011111010100	CTCACGTAGA	10	1036
C1127249	14	11010110011010010100	TAGACACTGA	10	1037
C1127249	14	11001010110101001100	TATCTAGATA	10	1038
C1127249	14	10100100110010110011	TCATATCGCG	10	1039
C1127249	14	10101001100110010110	TCTGTGTGAC	10	1040
C1127249	14	10010011101101011000	TGCGTCGAGT	10	1041

IDT Exhibit 1019.0256

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127249	14	10011011001001100110	TGTCGCACAC	10	1042
C1127250	14	01100110011001111000	ACACACACGT	10	1043
C1127250	14	01101010101101001010	ACTCTCGATC	10	1044
C1127250	14	01010011001110011001	AGCGCGTGTG	10	1045
C1127250	14	01001001011011010011	ATGACTAGCG	10	1046
C1127250	14	00110101010101001110	CGAGAGATAC	10	1047
C1127250	14	00111010010010010111	CGTCATGACG	10	1048
C1127250	14	00111001101010101100	CGTGTCTCTA	10	1049
C1127250	14	00101111101010010010	CTACGTCTGC	10	1050
C1127250	14	11100101001100101100	TACAGCGCTA	10	1051
C1127250	14	11011010010111001000	TAGTCAGTAT	10	1052
C1127250	14	10100110100111010100	TCACTGTAGA	10	1053
C1127250	14	10101001100100110011	TCTGTGCGCG	10	1054
C1127250	14	10010100111010011010	TGATACTGTC	10	1055
C1127250	14	10011011001001100110	TGTCGCACAC	10	1056
C1127251	14	01100110011001111000	ACACACACGT	10	1057
C1127251	14	01101010101101001010	ACTCTCGATC	10	1058
C1127251	14	01010011001110011001	AGCGCGTGTG	10	1059
C1127251	14	01001001011011010011	ATGACTAGCG	10	1060
C1127251	14	00110101010101001110	CGAGAGATAC	10	1061
C1127251	14	00111010010010010111	CGTCATGACG	10	1062
C1127251	14	00111001101010101100	CGTGTCTCTA	10	1063
C1127251	14	00101111101010010010	CTACGTCTGC	10	1064
C1127251	14	11100101001100101100	TACAGCGCTA	10	1065
C1127251	14	11011010010111001000	TAGTCAGTAT	10	1066
C1127251	14	10110100100111010100	TCGATGTAGA	10	1067
C1127251	14	10101001100100110011	TCTGTGCGCG	10	1068
C1127251	14	10010100111010011010	TGATACTGTC	10	1069
C1127251	14	10011011001001100110	TGTCGCACAC	10	1070
C1127252	14	01100110011001111000	ACACACACGT	10	1071
C1127252	14	01101010101101001010	ACTCTCGATC	10	1072
C1127252	14	01010011001110011001	AGCGCGTGTG	10	1073
C1127252	14	01001001011011010011	ATGACTAGCG	10	1074
C1127252	14	00110101010101001110	CGAGAGATAC	10	1075
C1127252	14	00111010010010010111	CGTCATGACG	10	1076
C1127252	14	00111001101010101100	CGTGTCTCTA	10	1077
C1127252	14	00101111101010010010	CTACGTCTGC	10	1078
C1127252	14	11100101001100101100	TACAGCGCTA	10	1079
C1127252	14	11011010010111001000	TAGTCAGTAT	10	1080
C1127252	14	10101100110011010100	TCTATATAGA	10	1081
C1127252	14	10101001100100110011	TCTGTGCGCG	10	1082
C1127252	14	10010100111010011010	TGATACTGTC	10	1083
C1127252	14	10011011001001100110	TGTCGCACAC	10	1084
C1127253	14	01100110011001111000	ACACACACGT	10	1085
C1127253	14	01101010101101001010	ACTCTCGATC	10	1086

IDT Exhibit 1019.0257

Cluster Id	Member Count	Flowgram TACGTACGTACGTACGTACG (SEQ ID NO: 6)	UID	UID Length	SEQ ID NO
C1127253	14	01010011001110011001	AGCGCGTGTG	10	1087
C1127253	14	01001001011011010011	ATGACTAGCG	10	1088
C1127253	14	00110101010101001110	CGAGAGATAC	10	1089
C1127253	14	00111010010010010111	CGTCATGACG	10	1090
C1127253	14	00111001101010101100	CGTGTCTCTA	10	1091
C1127253	14	00101111101010010010	CTACGTCTGC	10	1092
C1127253	14	11100101001100101100	TACAGCGCTA	10	1093
C1127253	14	11011010010111001000	TAGTCAGTAT	10	1094
C1127253	14	11001100110011010100	TATATATAGA	10	1095
C1127253	14	10101001100100110011	TCTGTGCGCG	10	1096
C1127253	14	10010100111010011010	TGATACTGTC	10	1097
C1127253	14	10011011001001100110	TGTCGCACAC	10	1098

Example 4- Exemplary computer code for representing and manipulating nucleotide sequences for UID identification

```
package com.fourfivefour.amplicons;

import java.util.HashSet;
import java.util.Set;

/**
 * Code to implement common operations on Nucleotide Sequences
 *
 *
 *
 */
public class Sequence implements Comparable<Sequence> {
        private String sequence;
        static final char possibleBases[] = { 'A', 'C', 'T', 'G' };

        public Sequence(String sequence) {
                this.sequence = sequence.toUpperCase();
        }

        public String getSequence() {
                return sequence;
        }

        public int hashCode() {
                return sequence.hashCode();
        }

        public boolean equals(Object obj) {
                return ((this == obj) ||
```

```
                ((obj instanceof Sequence) &&
                 sequence.equals(((Sequence) obj).sequence)));
        }

        public int compareTo(Sequence obj) {
                return sequence.compareTo(obj.sequence);
        }

        public String toString() {
                return sequence;
        }


        /**
         * Generate the set of all single base insertions for the
         * Sequence.
         *
         * @return          A set of Sequences representing all single base
         *          insertions of the Sequence.
         */
        public Set<Sequence> generateSingleInsertions() {
                Set<Sequence> insertions = new HashSet<Sequence>();

                int seqLen = sequence.length();
                for (int insertIdx = 0; insertIdx <= seqLen; insertIdx++) {
                        String prefixString = sequence.substring(0, insertIdx);
                        String suffixString = sequence.substring(insertIdx,seqLen);
                        for (char insertBase : possibleBases) {
                                insertions.add(new Sequence(prefixString + insertBase +
suffixString));
                        }
                }
                return insertions;
        }

        /**
         * Generate the set of all single base substitutions for the
         * Sequence.
         *
         * @return          A set of Sequences representing all single base
         *          substitutions of the Sequence.
         */
        public Set<Sequence> generateSingleSubstitutions() {
                Set<Sequence> substitutions = new HashSet<Sequence>();

                int seqLen = sequence.length();
                for (int substBaseIdx = 0; substBaseIdx < seqLen; substBaseIdx++) {
                        String prefixString =
                                sequence.substring(0, substBaseIdx);
                        String suffixString =
```

```
                        sequence.substring(substBaseIdx + 1, seqLen);
                    char originalBase =
                        sequence.charAt(substBaseIdx);
                    for (char substBase : possibleBases) {
                        if (substBase != originalBase) {
                            substitutions.add(
                             new Sequence(prefixString + substBase + suffixString)
                            );
                        }
                    }
                }
                return substitutions;
            }

            /**
             * Generate the set of all single base deletions for the
             * Sequence.
             *
             * @return              A set of sequences representing all single base
             *          deletions of the Sequence.
             */
            public Set<Sequence> generateSingleDeletions() {
                Set<Sequence> deletions = new HashSet<Sequence>();

                int seqLen = sequence.length();
                for (int deleteBaseIdx = 0; deleteBaseIdx < seqLen; deleteBaseIdx++) {
                    String prefixString =
                        sequence.substring(0, deleteBaseIdx);
                    String suffixString =
                        sequence.substring(deleteBaseIdx + 1, seqLen);
                    deletions.add(new Sequence(prefixString + suffixString));
                }
                return deletions;
            }

            /**
             * Generate all 1-base mutations starting from each of the sequences in
             * the input set of sequences.
             *
             * @param inputSeqs  The input set of sequences.
             * @return      A set of sequences that are exactly one mutation
             *          away from each of the sequences in the input set
             *          of sequences.
             */
            public static Set<Sequence> generateSingleMutations(Set<Sequence> inputSeqs) {
                Set<Sequence> mutatedSequences = new HashSet<Sequence>();
                for (Sequence inputSeq : inputSeqs) {
                    mutatedSequences.addAll(inputSeq.generateSingleDeletions());
                    mutatedSequences.addAll(inputSeq.generateSingleInsertions());
                    mutatedSequences.addAll(inputSeq.generateSingleSubstitutions());
```

```
            }
            return mutatedSequences;
        }
}
```

As stated previously, it will be appreciated that the foregoing computer code is provided for the purposes of example, and that numerous alternative methods and code structures may be employed. It will also be appreciated that the exemplary code provided herein is not intended to execute as a stand alone application or to run perfectly without additional computer code or modification.

Having described various embodiments and implementations, it should be apparent to those skilled in the relevant art that the foregoing is illustrative only and not limiting, having been presented by way of example only. Many other schemes for distributing functions among the various functional elements of the illustrated embodiment are possible. The functions of any element may be carried out in various ways in alternative embodiments.

IDT Exhibit 1019.0261

CLAIMS

What is claimed is:

1. An identifier element for identifying an origin of a template nucleic acid molecule, comprising:

a nucleic acid element comprising a sequence composition that enables detection of an introduced error in sequence data generated from the nucleic acid element and correction of the introduced error, wherein the nucleic acid element is constructed to couple with the end of a template nucleic acid molecule and identifies an origin of the template nucleic acid molecule.

2. The identifier element of claim 1, wherein:

the sequence composition enables detection of up to three of the introduced errors and correction for up to two of the introduced errors.

3. The identifier element of claim 1, wherein:

The sequence composition comprises 10 sequence positions.

4. The identifier element of claim 1, wherein:

the introduced error is selected from the group consisting of an insertion error, a deletion error, and a substitution error.

5. The identifier element of claim 1, wherein:

the sequence composition comprises a design based upon a set of parameters selected from the group consisting of minimum sequence length, minimum number of flow cycles, sequence distinctiveness, and monomer repeats.

6. The identifier element of claim 1, wherein:

the sequence composition comprises a design based upon a set of parameters selected from the group consisting of melting temperature, Gibbs free energy, hairpin formation, and dimer formation.

7. The identifier element of claim 1, wherein:

IDT Exhibit 1019.0262

the nucleic acid element is incorporated into an adaptor comprising a primer element, wherein the adaptor couples with the end of the template nucleic acid molecule.

8. The identifier element of claim 7, wherein:
the nucleic acid element is in a known position relative to the primer element.

9. The identifier element of claim 7, wherein:
the primer element is selected from the group consisting of an amplification primer, a sequencing primer, or a bipartite amplification – sequencing primer.

10. The identifier element of claim 7, wherein:
the adaptor comprises a quality control element.

11. The identifier element of claim 7, wherein:
the nucleic acid element is in a known position relative to the quality control element.

12. The identifier element of claim 1, wherein:
the origin of the template nucleic acid molecule comprises an experimental sample or diagnostic sample.

13. The identifier element of claim 1, wherein:
the nucleic acid element belongs to a set comprising a plurality of compatible nucleic acid elements each comprising a distinctive sequence composition, wherein the detection of the introduced error is relative to the sequence composition of the compatible nucleic acid elements of the set.

14. The identifier element of claim 13, wherein:
the set comprises 14 of the compatible nucleic acid elements.

15. A method for identifying an origin of a template nucleic acid molecule, comprising the steps of:
identifying a first identifier sequence from sequence data generated from a template nucleic acid molecule;

IDT Exhibit 1019.0263

detecting an introduced error in the first identifier sequence;

correcting the introduced error in the first identifier sequence;

associating the corrected first identifier sequence with a first identifier element coupled to the template molecule; and

identifying an origin of the template molecule using the association of the corrected first identifier sequence with the first identifier element.

16. The method of claim 15, further comprising:

sequencing a template nucleic acid molecule to generate the sequence data.

17. The method of claim 15, wherein:

the template nucleic acid molecule is included in a multiplex sample comprising a plurality of template molecules from a plurality of different origins.

18. The method of claim 15, further comprising:

detecting up to three of the introduced errors in the first identifier sequence; and

correcting up to two of the introduced errors in the first identifier sequence.

19. The method of claim 15, wherein:

the introduced error is selected from the group consisting of an insertion error, a deletion error, and a substitution error.

20. The method of claim 15, wherein the step of detecting comprises:

measuring one or more characteristics of sequence composition in one or more sequence regions that flank the identifier sequence; and

detecting the introduced error using one or more assumptions derived from the measured characteristics.

21. The method of claim 15, wherein:

the first identifier element is incorporated into an adaptor comprising a primer element, wherein the adaptor is coupled to the template nucleic acid molecule.

22. The method of claim 21, wherein:

IDT Exhibit 1019.0264

the first identifier element is in a known position relative to the primer element.

23. The method of claim 21, wherein:

the primer element is selected from the group consisting of an amplification primer, a sequencing primer, or a bipartite amplification – sequencing primer.

24. The method of claim 21, wherein:

the adaptor comprises a quality control element.

25. The method of claim 21, wherein:

the first identifier element is in a known position relative to the quality control element.

26. The method of claim 15, wherein:

the origin of the template nucleic acid molecule comprises an experimental sample or diagnostic sample.

27. The method of claim 15, further comprising the steps of:

identifying a second identifier sequence from the sequence data generated from the template nucleic acid molecule;

detecting an introduced error in the second identifier sequence;

correcting the introduced error in the second identifier sequence;

associating the corrected second identifier sequence with a second identifier element coupled with the template nucleic acid molecule; and

identifying an origin of the template nucleic acid molecule using the association of the corrected second identifier sequence with the second identifier element combinatorially with the association of the corrected first identifier sequence with the first identifier element.

28. The method of claim 27, further comprising:

detecting up to three of the introduced errors in the second identifier sequence; and

correcting up to two of the introduced errors in the second identifier sequence.

29. The method of claim 15, wherein:

IDT Exhibit 1019.0265

the introduced error is selected from the group consisting of an insertion error, a deletion error, and a substitution error.

30. The method of claim 15, wherein:

the first identifier belongs to at least one set of compatible identifiers of a plurality of sets of identifiers.

31. The method of claim 15, wherein:

the set of compatible identifiers comprise 14 identifiers that enable the detection and the correction of the introduced error.

32. A kit for identifying an origin of a template nucleic acid molecule comprising:

a set of nucleic acid elements each comprising a distinctive sequence composition that enables detection of an introduced error in sequence data generated from each nucleic acid element and correction of the introduced error, wherein each of the nucleic acid elements is constructed to couple with the end of a template nucleic acid molecule and identifies the origin of the template nucleic acid molecule.

34. The kit of claim 32, wherein:

the distinctive sequence composition enables detection of up to three of the introduced errors and correction for up to two of the introduced errors.

35. The kit of claim 32, wherein:

the introduced error is selected from the group consisting of an insertion error, a deletion error, and a substitution error.

36. The kit of claim 32, wherein:

each nucleic acid element is incorporated into an adaptor comprising a primer element, wherein the adaptor couples with the end of the template nucleic acid molecule.

37. The kit of claim 36, wherein:

the nucleic acid element is in a known position relative to the primer element.

IDT Exhibit 1019.0266

38. The kit of claim 36, wherein:

the primer element is selected from the group consisting of an amplification primer, a sequencing primer, or a bipartite amplification – sequencing primer.

39. The kit of claim 36, wherein:

the adaptor comprises a quality control element.

40. The kit of claim 36, wherein:

the nucleic acid element is in a known position relative to the quality control element.

41. The kit of claim 32, wherein:

the detection of the introduced error in each of the nucleic acid elements is relative to the distinctive sequence composition of the other nucleic acid elements of the set.

42. The kit of claim 41, wherein:

the set comprises 14 of the nucleic acid elements.

43. A computer, comprising executable code stored thereon, wherein the executable code performs a method for identifying an origin of a template nucleic acid molecule, comprising the steps of:

identifying an identifier sequence from sequence data generated from a template nucleic acid molecule;

detecting an introduced error in the identifier sequence;

correcting the introduced error in the identifier sequence;

associating the corrected identifier sequence with an identifier element coupled with the template molecule; and

identifying an origin of the template molecule using the association of the corrected identifier sequence with the identifier element.

44. The method of claim 43, wherein:

the template nucleic acid molecule is included in a multiplex sample comprising a plurality of template molecules from a plurality of different origins.

IDT Exhibit 1019.0267

45. The method of claim 43, further comprising:

detecting up to three of the introduced errors in the first identifier sequence; and

correcting up to two of the introduced errors in the first identifier sequence.

46. The method of claim 43, wherein:

the introduced error is selected from the group consisting of an insertion error, a deletion error, and a substitution error.

48. The method of claim 43, wherein the step of identifying further comprises:

determining a position for the identifier sequence using a known positional relationship of one or more elements in the sequence data.

49. The method of claim 48, wherein:

the one or more elements include a primer sequence.

50. The method of claim 43, wherein the step of detecting further comprises:

measuring one or more characteristics of sequence composition in one or more sequence regions that flank the identifier sequence; and

detecting the introduced error using one or more assumptions derived from the measured characteristics.

51. The method of claim 43, further comprising:

identifying a second identifier sequence from the sequence data generated from the template nucleic acid molecule;

detecting an introduced error in the second identifier sequence;

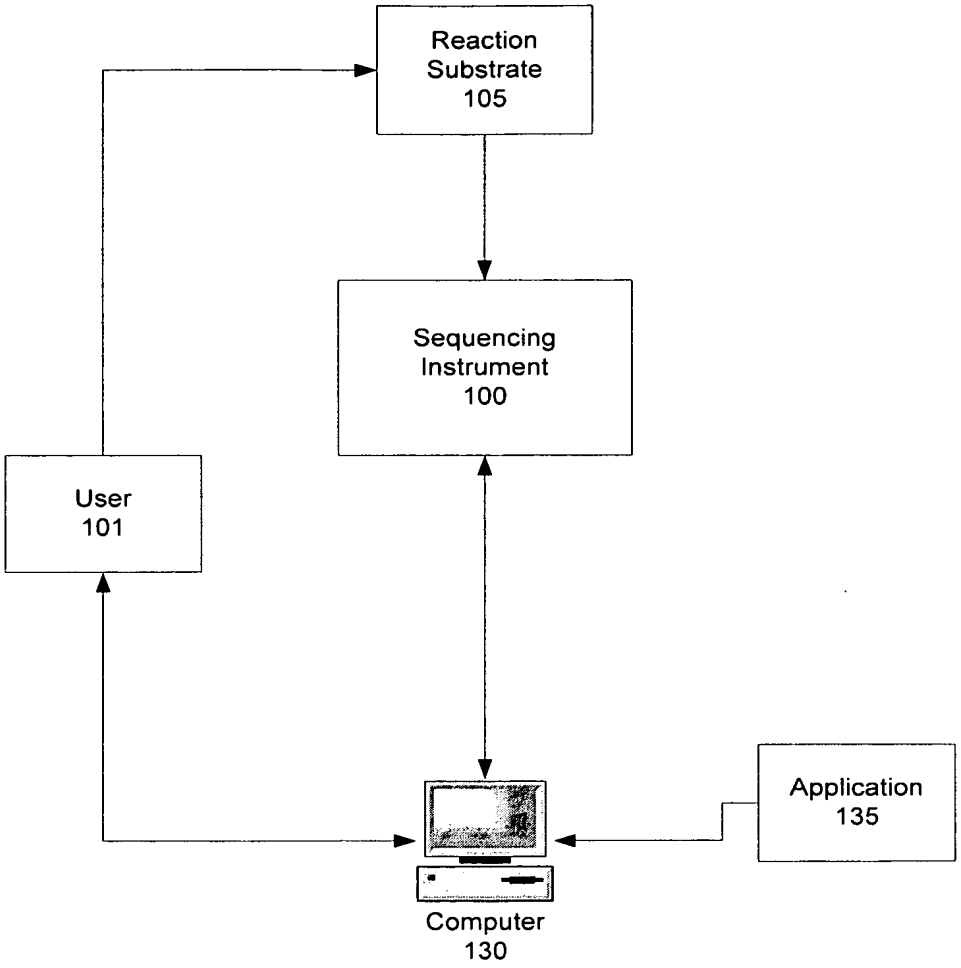
correcting the introduced error in the second identifier sequence;

associating the corrected second identifier sequence with a second identifier element coupled with the template molecule; and

identifying an origin of the template molecule using the association of the corrected second identifier sequence with the second identifier element combinatorially with the association of the corrected first identifier sequence with the first identifier element.

IDT Exhibit 1019.0268

FIGURE 1



IDT Exhibit 1019.0269

FIGURE 2A

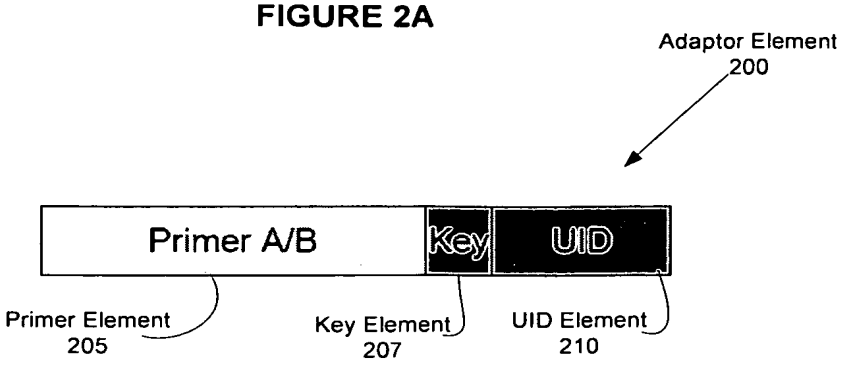


FIGURE 2B

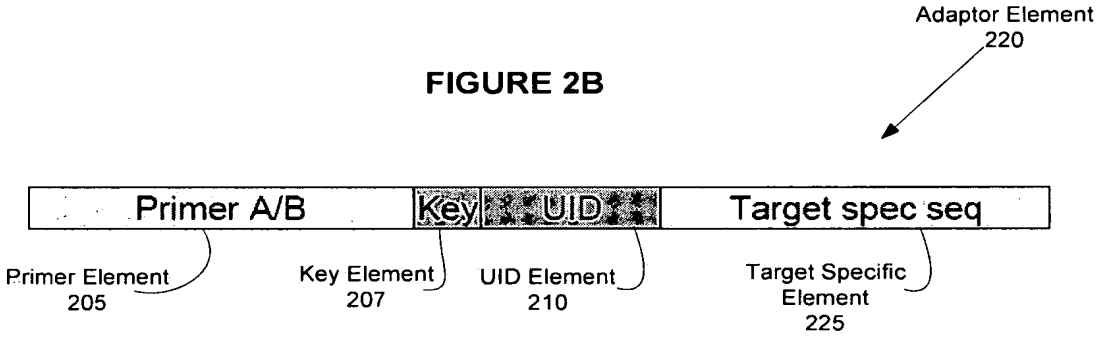
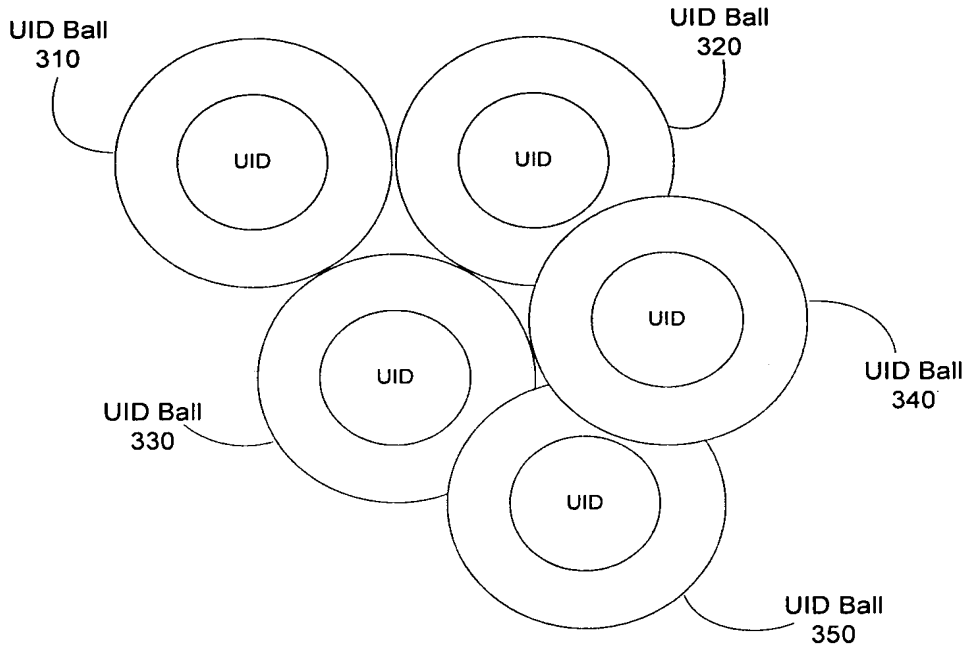


FIGURE 3


UID Ball
310
UID Ball
320
UID
UID
UID
UID
UID
UID Ball
340
UID Ball
330
UID Ball
350

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 08/06822

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68 (2008.04)
USPC - 435/6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC - 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/91.2; 536/23.1; 536/25.32; 536/25.4; 536/26.6

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST(PGPB,USPT,USOC,EPAB,JPAB), Google Scholar: (sequencing adaptors origin identify error, "Patrick S. Schnable" origin, "Daniel A. Ashlock" algorithm, sequencing adaptors quality control element error, sequencing adaptors "quality control" element error, sequencing adaptors primer "quality control element" error)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	QIU et al. DNA Sequence-Based "Bar Codes" for Tracking the Origins of Expressed Sequence Tags from a Maize cDNA Library Constructed Using Multiple mRNA Sources. Plant Physiology, 2003, vol 133, pp 475-481; Abstract; pg 475, col 2 para 3; pg 476, col 1 para 1, ln 1-8, para 3; col 2 para 1, ln 1-2; pg 480, col 1, para 3, ln 5-11,14-17, col 2 para 2, ln 7-13, para 3, 6, ln 2-3, para 7, ln 4-5; pg 481, col 1, para 2, 4, 8, ln 1-6	1-8, 12-22, 26-31, 43-46, 48-51 --- 9-11, 23-25, 32, 34-42
Y	US 2007/00772208 A1 (DRMANAC) 29 Mar 2007 (29.03.2007); para [0015], [0053]	9-11, 23-25, 32, 34-42
Y	US 2005/0221341 A1 (SHIMKETS et al) 6 Oct 2005 (06.10.2005); para [0050], [0053]	10-11, 24-25, 39-40

☐ 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
01 September 2008 (01.09.2008)	08 SEP 2008
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: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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					

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	2013/190441	WO	A2	2013-12-27	Univ Hong Kong Chinese		
	2	2013/191775	WO	A2	2013-12-27	Nugen Technologies, Inc		
	3	2014/039556	WO	A1	2014-03-13	Guardant Health, Inc		

IDT Exhibit 1019.0273

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	4	2014/082032	WO	A1	2014-05-30	Univ Toledo		
	5	2014/144092	WO	A1	2014-09-18	Nugen Technologies, Inc		
	6	2014/150931	WO	A1	2014-09-25	Nugen Technologies, Inc		
	7	2015/031691	WO	A1	2015-03-05	Cellular Research, Inc		
	8	2015/073711	WO	A1	2015-05-21	Nugen Technologies, Inc		
	9	2015/131107	WO	A1	2015-09-03	Nugen Technologies, Inc		
	10	89/09284	WO	A1	1989-10-05	Univ Iowa Res Found		
	11	92/07951	WO	A1	1992-05-14	Kow, Yoke		
	12	93/18052	WO	A1	1993-09-16	Isis Pharmaceuticals, Inc		
	13	94/16090	WO	A1	1994-07-21	Molecular Tool, Inc		
	14	96/40998	WO	A1	1996-12-19	Pioneer Hi-bred International, Inc		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	15	97/12061	WO	A1	1997-04-03	Epicentre Technologies Corp		
	16	97/25416	WO	A2	1997-07-17	Nyfotek As		
	17	98/044151	WO	A1	1998-10-08	Glaxo Group Ltd		
	18	98/06736	WO	A1	1998-02-19	Life Technologies, Inc		
	19	98/38296	WO	A1	1998-09-03	Gesher Israel Advanced Biotecs		
	20	99/10540	WO	A1	1999-03-04	Lopez, Osvaldo		
	21	99/11819	WO	A1	1999-03-11	Perkin Elmer Corp		
	22	99/42618	WO	A1	1999-08-26	Dade Behring Inc		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

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		

If you wish to add additional non-patent literature document citation information please click the Add button Add

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2020-07-22
Name/Print	Thomas C. Meyers	Registration Number	36,989

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.

IDT Exhibit 1019.0278

Electronic Acknowledgement Receipt

EFS ID:	40069748
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Kelley Warren
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	22-JUL-2020
Filing Date:	15-JAN-2020
Time Stamp:	15:40:19
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	Transmittal Letter	Information_Disclosure_Statement.pdf	21398 2d4b2e09eff92df444ac82fd29b0a9b5e25c4d49	no	2

Warnings:

Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	34079_516US_-_Information_Disclosure_Statement_Fillable_PDF_1_of_3.pdf	1046427 4315a7912fa72ee03b2b68e16950e6dfa635a8d4	no	40
Warnings:					
Information:					
3	Information Disclosure Statement (IDS) Form (SB08)	34079_516US_-_Information_Disclosure_Statement_Fillable_PDF_2_of_3.pdf	1038479 865647fa06270a7f6a734ce63410d0ec087d033a	no	17
Warnings:					
Information:					
4	Information Disclosure Statement (IDS) Form (SB08)	34079_516US_-_Information_Disclosure_Statement_Fillable_PDF_3_of_3.pdf	1035112 3f51cc7caa4bfd0bc95d63677141b3607aa0231f	no	6
Warnings:					
Information:					
A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.					
5	Foreign Reference	34079_12US_-_WO2008150432.pdf	5241093 b677cbabd678730feea1a307154627dd261f25f6	no	93
Warnings:					
Information:					
			Total Files Size (in bytes):	8382509	

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.

Docket No.: NUGN-001/03US
34079/516
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
NUGEN TECHNOLOGIES, INC.

Application No.: 16/743,604 Confirmation No.: 3249

Filed: January 15, 2020 Art Unit: 1637

For: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ Examiner: T. E. Strzelecka

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

INFORMATION DISCLOSURE STATEMENT (IDS)

Pursuant to 37 C.F.R. §§ 1.56, 1.97, and 1.98, the attention of the United States Patent and Trademark Office is hereby directed to the references listed on the attached PTO/SB/08. It is respectfully requested that the information be expressly considered during the prosecution of the above-identified application, and that the references be made of record therein and appear among the "References Cited" on any patent to issue therefrom.

Copies of the references listed below are attached:

For. Doc. No.	Ctry	Applicant/Patentee	Publication Date
2008150432	WO	454 Life Sciences Corp	December 11, 2008

In accordance with 37 C.F.R. § 1.98(d)(1) the remaining references are not supplied because they were previously cited by or submitted to the Office in prior application numbers 15/406,002 filed January 13, 2017 and 14/540,917 filed November 13, 2014 and relied on in the above-identified application for an earlier effective filing date under 35 U.S.C. § 120.

63815944

It is submitted that the Information Disclosure Statement is in compliance with 37 C.F.R. § 1.98, and the Examiner is respectfully requested to consider the listed references.

Applicant believes no fee is due with this response.

Dated: July 22, 2020

Respectfully submitted,

Electronic signature: /Thomas C. Meyers/
Thomas C. Meyers
Registration No.: 36,989
BROWN RUDNICK LLP
Brown Rudnick LLP
One Financial Center
Boston, Massachusetts 02111
(617) 856-8200
Attorney for Applicant

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	20090029385	A1	2009-01-29	Christians et al.	
	2	20090036663	A1	2009-02-05	Kurn	
	3	20090061425	A1	2009-03-05	Lo et al.	
	4	20090061439	A1	2009-03-05	Buzby	
	5	20090068645	A1	2009-03-12	Sibson	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	6	20090068655	A1	2009-03-12	Williams	
	7	20090068709	A1	2009-03-12	Kurn et al.	
	8	20090105081	A1	2009-04-23	Rodesch et al.	
	9	20090117573	A1	2009-05-07	Fu et al.	
	10	20090117621	A1	2009-05-07	Boutell et al.	
	11	20090124514	A1	2009-05-14	Fu et al.	
	12	20090127589	A1	2009-05-21	Rothberg et al.	
	13	20090130721	A1	2009-05-21	Kurn et al.	
	14	20090203085	A1	2009-08-13	KURN et al.	
	15	20090203531	A1	2009-08-13	Kurn	
	16	20090233802	A1	2009-09-17	Bignell et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	17	20090233804	A1	2009-09-17	Kurn et al.	
	18	20090239232	A1	2009-09-24	Kurn	
	19	20090275486	A1	2009-11-05	Kurn et al.	
	20	20090280538	A1	2009-11-12	Patel et al.	
	21	20090298075	A1	2009-12-03	Travers et al.	
	22	20100015666	A1	2010-01-21	Brenner et al.	
	23	20100021973	A1	2010-01-28	Makarov et al.	
	24	20100022403	A1	2010-01-28	Kurn et al.	
	25	20100029511	A1	2010-02-04	Raymond et al.	
	26	20100105052	A1	2010-04-29	DRMANAC et al.	
	27	20100113296	A1	2010-05-06	Myerson	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	28	20100129879	A1	2010-05-27	Ach et al.	
	29	20100137143	A1	2010-06-03	Rothberg et al.	
	30	20100159559	A1	2010-06-24	Kurn et al.	
	31	20100167954	A1	2010-07-01	Earnshaw et al.	
	32	20100173394	A1	2010-07-08	Colston, JR. et al.	
	33	20100203597	A1	2010-08-12	Chen et al.	
	34	20100267043	A1	2010-10-21	Braverman et al.	
	35	20100273219	A1	2010-10-28	May et al.	
	36	20100311066	A1	2010-12-09	Kurn	
	37	20100323348	A1	2010-12-23	Hamady et al.	
	38	20110015096	A1	2011-01-20	CHIU	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	39	20110039732	A1	2011-02-17	Raymond et al.	
	40	20110104785	A1	2011-05-05	Vaidyanathan et al.	
	41	20110105364	A1	2011-05-05	Kurn	
	42	20110129827	A1	2011-06-02	Causey et al.	
	43	20110189679	A1	2011-08-04	Kurn et al.	
	44	20110224105	A1	2011-09-15	Kurn et al.	
	45	20110288780	A1	2011-11-24	Rabinowitz et al.	
	46	20110294132	A1	2011-12-01	Kurn	
	47	20110319290	A1	2011-12-29	Raymond et al.	
	48	20120003657	A1	2012-01-05	Myllykangas et al.	
	49	20120028310	A1	2012-02-02	Kurn et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	50	20120045797	A1	2012-02-23	Kurn et al.	
	51	20120071331	A1	2012-03-22	Casbon et al.	
	52	20120074925	A1	2012-03-29	Oliver	
	53	20120102054	A1	2012-04-26	POPESCU et al.	
	54	20120107811	A1	2012-05-03	Kelso et al.	
	55	20120122701	A1	2012-05-17	Ryan et al.	
	56	20120149068	A1	2012-06-14	Kurn	
	57	20120156728	A1	2012-06-21	Li et al.	
	58	20120157322	A1	2012-06-21	Myllykangas et al.	
	59	20120190587	A1	2012-07-26	Kurn et al.	
	60	20120208705	A1	2012-08-16	Steemers et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	61	20120220483	A1	2012-08-30	KURN et al.	
	62	20120220494	A1	2012-08-30	Samuels et al.	
	63	20120237943	A1	2012-09-20	Soldatov et al.	
	64	20120238738	A1	2012-09-20	Hendrickson	
	65	20120245041	A1	2012-09-27	Brenner et al.	
	66	20120270212	A1	2012-10-25	RABINOWITZ et al.	
	67	20120283145	A1	2012-11-08	Wang	
	68	20120289426	A1	2012-11-15	Roos et al.	
	69	20120309002	A1	2012-12-06	Link	
	70	20130005585	A1	2013-01-03	Anderson et al.	
	71	20130059738	A1	2013-03-07	LEAMON et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	72	20130231253	A1	2013-09-05	Amorese et al.	
	73	20140038188	A1	2014-02-06	Kurn	
	74	20140038236	A1	2014-02-06	Kurn et al.	
	75	20140065692	A1	2014-03-06	KURN et al.	
	76	20140274729	A1	2014-09-18	Kurn et al.	
	77	20140274731	A1	2014-09-18	RAYMOND et al.	
	78	20140274738	A1	2014-09-18	Amorese et al.	
	79	20140303000	A1	2014-10-09	Armour	
	80	20140378345	A1	2014-12-25	Hindson et al.	
	81	20150011396	A1	2015-01-08	Schroeder et al.	
	82	20150017635	A1	2015-01-15	Myllykangas et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	83	20150101595	A1	2015-04-16	Hancock et al.	
	84	20150133319	A1	2015-05-14	Fu et al.	
	85	20150284769	A1	2015-10-08	Schroeder	
	86	20150299767	A1	2015-10-22	Armour et al.	
	87	20150299784	A1	2015-10-22	Fan et al.	
	88	20150299812	A1	2015-10-22	Talasaz	
	89	20160122756	A1	2016-05-05	Armour	
	90	20160130576	A1	2016-05-12	Armour	
	91	20160153039	A1	2016-06-02	Amorese et al.	
	92	20160203259	A1	2016-07-14	SCOLNICK et al.	
	93	20160220994	A1	2016-08-04	Wright	

IDT Exhibit 1019.0292

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	94	20160251711	A1	2016-09-01	AMORESE et al.	
	95	20160265042	A1	2016-09-15	Schroeder et al.	
	96	20160275240	A1	2016-09-22	HUELGA et al.	
	97	20160296930	A1	2016-10-13	Matear 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	2006/086668	WO	A2	2006-08-17	Epict Technologies		
	2	2006/137733	WO	A1	2006-12-28	Keygene NV		
	3	2007/018601	WO	A1	2007-02-15	Rubicon Genomics, Inc		
	4	2007/019444	WO	A2	2007-02-15	Euclid Diagnostics LLC		
	5	2007/030759	WO	A2	2007-03-15	Nugen Technologies, Inc		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	6	2007/037678	WO	A2	2007-04-05	Keygene NV		
	7	2007/052006	WO	A1	2007-05-10	Solexa Ltd		
	8	2007/057652	WO	A1	2007-05-24	Solexa Ltd		
	9	2007/073165	WO	A1	2007-06-28	Keygene NV		
	10	2007/136717	WO	A1	2007-11-29	Nugen Technologies, Inc		
	11	2008/005459	WO	A2	2008-01-10	Nugen Technologies, Inc		
	12	2008/015396	WO	A2	2008-02-07	Solexa Ltd		
	13	2008/033442	WO	A2	2008-03-20	Univ Leland Stanford Junior		
	14	2008/093098	WO	A2	2008-08-07	Solexa Ltd		
	15	2008/115185	WO	A2	2008-09-25	Nimblegen Systems, Inc		
	16	2008150432	WO	A1	2008-12-11	454 Life Sciences Corp		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	17	2009/053039	WO	A1	2009-04-30	Roche Diagnostics GmbH		
	18	2009/102878	WO	A2	2009-08-20	Nugen Technologies, Inc		
	19	2009/102896	WO	A2	2009-08-20	Nugen Technologies, Inc		
	20	2009/112844	WO	A1	2009-09-17	Stichting Het Nl Kanker I		
	21	2009/117698	WO	A2	2009-09-24	Nugen Technologies, Inc		
	22	2009/120372	WO	A2	2009-10-01	Pacific Biosciences Of California, Inc		
	23	2009/120374	WO	A2	2009-10-01	Pacific Biosciences Of California, Inc		
	24	2010/003153	WO	A2	2010-01-07	Life Technologies Corp		
	25	2010/030683	WO	A1	2010-03-18	Rosetta Inpharmatics LLC		
	26	2010/039991	WO	A2	2010-04-08	Texas A & M Univ Sys		
	27	2010/063711	WO	A1	2010-06-10	Angeletti P Ist Richerche Bio		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	28	2010/064893	WO	A1	2010-06-10	Keygene NV		
	29	2010/085715	WO	A1	2010-07-29	Quanta Biosciences		
	30	2010/115154	WO	A1	2010-10-07	Fluidigm Corp		
	31	2010/129937	WO	A2	2010-11-11	Life Technologies Corp		
	32	2010091246	WO	A2	2010-08-12	Univ Northwestern		
	33	2011/003630	WO	A1	2011-01-13	Max Planck Gesellschaft		
	34	2011/009941	WO	A1	2011-01-27	Olink Genomics AB		
	35	2011/019964	WO	A1	2011-02-17	Nugen Technologies, Inc		
	36	2011/032053	WO	A1	2011-03-17	Nugen Technologies, Inc		
	37	2011/053987	WO	A1	2011-05-05	Nugen Technologies, Inc		
	38	2011/151777	WO	A1	2011-12-08	Consiglio Nazionale Ricerche		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	39	2011/156529	WO	A2	2011-12-15	Nugen Technologies, Inc		
	40	2011032040	WO	A1	2011-03-17	Centrillion Technology Holding Corp		
	41	2012/013932	WO	A1	2012-02-02	Univ St Andrews		
	42	2012/054873	WO	A3	2012-08-23	Cold Spring Harbor Lab		
	43	2012/061832	WO	A1	2012-05-10	Illumina, Inc		
	44	2012/103154	WO	A1	2012-08-02	Nugen Technologies, Inc		
	45	2013/059740	WO	A1	2013-04-25	Foundation Medicine, Inc		
	46	2013/059746	WO	A1	2013-04-25	Nugen Technologies, Inc		
	47	2013/112923	WO	A1	2013-08-01	Nugen Technologies, Inc		
	48	2013/130512	WO	A3	2013-10-24	Univ North Carolina		
	49	2013/138510	WO	A9	2014-07-10	Patel, Abhijit Ajit		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	50	2013/177220	WO	A1	2013-11-28	Scripps Research Inst		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

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	Westin 2000, Anchored multiplex amplification on a microelectronic chip array, Nat Biotech 18:199-204	
	2	Wienholds, 2004, Target-selected gene inactivation in zebrafish, Meth Cell Biol 77:69-90	
	3	Wolford, 2000, High-throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC), Hum Genet 107:483-487	
	4	Xi, 2011, Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion, PNAS 108(46):e1128-e1136	
	5	Xu, 2012, FastUniq: A fast de novo duplicates removal tool for paired short reads, PLoSOne 7(12):e52249	

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2020-07-22
Name/Print	Thomas C. Meyers	Registration Number	36,989

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.**

IDT Exhibit 1019.0299

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.

IDT Exhibit 1019.0300

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	4362867		1982-12-07	Paddock	
	2	4458066		1984-07-03	Caruthers et al.	
	3	4469863		1984-09-04	Ts'o et al.	
	4	4582877		1986-04-15	Fairchok et al.	
	5	4876187		1989-10-24	Duck et al.	
	6	4925065		1990-05-15	Golias	
	7	4935357		1990-06-19	Szybalski	
	8	4942124		1990-07-17	Church	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	9	4988617		1991-01-29	Landegren et al.	
	10	4996143		1991-02-26	Heller et al.	
	11	5011769		1991-04-30	Duck et al.	
	12	5034506		1991-07-23	Summerton et al.	
	13	5035996		1991-07-30	Hartley	
	14	5043272		1991-08-27	Hartley	
	15	5082830		1992-01-21	Brakel et al.	
	16	5090591		1992-02-25	Long	
	17	5130238		1992-07-14	Malek et al.	
	18	5169766		1992-12-08	Schuster et al.	
	19	5171534		1992-12-15	Smith et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	20	5194370		1993-03-16	Berninger et al.	
	21	5216141		1993-06-01	Benner	
	22	5234809		1993-08-10	Boom et al.	
	23	5235033		1993-08-10	Summerton et al.	
	24	5242794		1993-09-07	Whiteley et al.	
	25	5384242		1995-01-24	Oakes	
	26	5386023		1995-01-31	Sanghvi et al.	
	27	5399491		1995-03-21	Kacian et al.	
	28	5409818		1995-04-25	Davey et al.	
	29	5418149		1995-05-23	Gelfand et al.	
	30	5422271		1995-06-06	Chen et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	31	5427929		1995-06-27	Richards et al.	
	32	5480784		1996-01-02	Kacian et al.	
	33	5494810		1996-02-27	Barany et al.	
	34	5508169		1996-04-16	Deugau et al.	
	35	5508178		1996-04-16	Rose et al.	
	36	5510270		1996-04-23	Fodor et al.	
	37	5525471		1996-06-11	Zeng	
	38	5545522		1996-08-13	Van Gelder et al.	
	39	5554516		1996-09-10	Kacian et al.	
	40	5554517		1996-09-10	Davey et al.	
	41	5556752		1996-09-17	Lockhart et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	42	5565340		1996-10-15	Chenchik et al.	
	43	5573913		1996-11-12	Rosemeyer et al.	
	44	5578832		1996-11-26	Trulson et al.	
	45	5589339		1996-12-31	Hampson et al.	
	46	5602240		1997-02-11	De Mesmaeker et al.	
	47	5637684		1997-06-10	Cook et al.	
	48	5641658		1997-06-24	Adams et al.	
	49	5644048		1997-07-01	Yau	
	50	5665549		1997-09-09	Pinkel et al.	
	51	5667976		1997-09-16	Van Ness et al.	
	52	5667979		1997-09-16	Berrens	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	53	5679512		1997-10-21	Laney et al.	
	54	5681726		1997-10-28	Huse et al.	
	55	5683879		1997-11-04	Laney et al.	
	56	5688648		1997-11-18	Mathies et al.	
	57	5705628		1998-01-06	Hawkins	
	58	5708154		1998-01-13	Smith et al.	
	59	5710028		1998-01-20	Eyal et al.	
	60	5712126		1998-01-27	Weissman et al.	
	61	5716785		1998-02-10	Van Gelder et al.	
	62	5726329		1998-03-10	Jones et al.	
	63	5750341		1998-05-12	Macevicz	

IDT Exhibit 1019.0306

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	64	5759822		1998-06-02	Chenchik et al.	
	65	5763178		1998-06-09	Chirikjian et al.	
	66	5789206		1998-08-04	Tavtigian et al.	
	67	5824517		1998-10-20	Cleuziat et al.	
	68	5824518		1998-10-20	Kacian et al.	
	69	5837832		1998-11-17	Chee et al.	
	70	5876976		1999-03-02	Richards et al.	
	71	5882867		1999-03-16	Ullman et al.	
	72	5888779		1999-03-30	Kacian et al.	
	73	5888819		1999-03-30	Goelet et al.	
	74	5945313		1999-08-31	Hartley et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	75	5952176		1999-09-14	McCarthy et al.	
	76	5958681		1999-09-28	Wetmur et al.	
	77	5965409		1999-10-12	Pardee et al.	
	78	5969119		1999-10-19	Macevicz	
	79	5972618		1999-10-26	Bloch	
	80	6004744		1999-12-21	Goelet et al.	
	81	6004745		1999-12-21	Arnold, Jr. et al.	
	82	6027889		2000-02-22	Barany et al.	
	83	6027923		2000-02-22	Wallace	
	84	6030774		2000-02-29	Laney et al.	
	85	6037152		2000-03-14	Richards et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	86	6056661		2000-05-02	Schmidt	
	87	6077674		2000-06-20	Schleifer et al.	
	88	6087103		2000-07-11	Burmer	
	89	6090553		2000-07-18	Matson	
	90	6090591		2000-07-18	Burg et al.	
	91	6107023		2000-08-22	Reyes et al.	
	92	6110709		2000-08-29	Ausubel et al.	
	93	6150112		2000-11-21	Weissman et al.	
	94	6159685		2000-12-12	Pinkel et al.	
	95	6160105		2000-12-12	Cunningham et al.	
	96	6169194		2001-01-02	Thompson et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	97	6172208		2001-01-09	Cook	
	98	6174680		2001-01-16	Makrigiorgos	
	99	6180338		2001-01-30	Adams	
	100	6190865		2001-02-20	Jendrisak et al.	
	101	6194211		2001-02-27	Richards et al.	
	102	6197501		2001-03-06	Cremer et al.	
	103	6197557		2001-03-06	Makarov et al.	
	104	6210891		2001-04-03	Nyren et al.	
	105	6225109		2001-05-01	Juncosa et al.	
	106	6225451		2001-05-01	Ballinger et al.	
	107	6232104		2001-05-15	Lishanski et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	108	6251639		2001-06-26	Kurn	
	109	6262490		2001-07-17	Hsu et al.	
	110	6270961		2001-08-07	Drmanac	
	111	6280935		2001-08-28	Macevicz	
	112	6287766		2001-09-11	Nolan et al.	
	113	6287825		2001-09-11	Weissman et al.	
	114	6291170		2001-09-18	Van Gelder et al.	
	115	6306365		2001-10-23	Ruoslahti et al.	
	116	6306597		2001-10-23	Macevicz	
	117	6309843		2001-10-30	Timms	
	118	6326142		2001-12-04	Royer	

IDT Exhibit 1019.0311

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	119	6335167		2002-01-01	Pinkel et al.	
	120	6339147		2002-01-15	Lukhtanov et al.	
	121	6440705		2002-08-27	Stanton, Jr. et al.	
	122	6449562		2002-09-10	Chandler et al.	
	123	6582938		2003-06-24	Su et al.	
	124	6670461		2003-12-30	Wengel et al.	
	125	6686156		2004-02-03	Kurn	
	126	6692918		2004-02-17	Kurn	
	127	6770748		2004-08-03	Imanishi et al.	
	128	6777180		2004-08-17	Fisher et al.	
	129	6794499		2004-09-21	Wengel et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	130	6815164		2004-11-09	Kurn	
	131	6815167		2004-11-09	Crothers et al.	
	132	6825011		2004-11-30	Romantchikov	
	133	6833246		2004-12-21	Balasubramanian	
	134	6849404		2005-02-01	Park et al.	
	135	6858413		2005-02-22	Kurn	
	136	6913884		2005-07-05	Stuelpnagel et al.	
	137	6917726		2005-07-12	Levene et al.	
	138	6924104		2005-08-02	Weissman et al.	
	139	6946251		2005-09-20	Kurn	
	140	7001724		2006-02-21	Greenfield	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	141	7033764		2006-04-25	Korlach et al.	
	142	7048481		2006-05-23	Sugata et al.	
	143	7052847		2006-05-30	Korlach et al.	
	144	7056676		2006-06-06	Korlach et al.	
	145	7056716		2006-06-06	Potter et al.	
	146	7060441		2006-06-13	Bourget et al.	
	147	7094536		2006-08-22	Kurn	
	148	7115400		2006-10-03	Adessi et al.	
	149	7170050		2007-01-30	Turner et al.	
	150	7175982		2007-02-13	McCarthy et al.	
	151	7176025		2007-02-13	Kurn et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	152	7189512		2007-03-13	Porat et al.	
	153	7211390		2007-05-01	Rothberg et al.	
	154	7232656		2007-06-19	Balasubramanian et al.	
	155	7244559		2007-07-17	Rothberg et al.	
	156	7244567		2007-07-17	Chen et al.	
	157	7264929		2007-09-04	Rothberg et al.	
	158	7273730		2007-09-25	Du Breuil Lastrucci	
	159	7276720		2007-10-02	Ulmer	
	160	7294461		2007-11-13	Kurn	
	161	7300755		2007-11-27	Petersdorf et al.	
	162	7302146		2007-11-27	Turner et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	163	7313308		2007-12-25	Turner et al.	
	164	7315019		2008-01-01	Turner et al.	
	165	7323305		2008-01-29	Leamon et al.	
	166	7335762		2008-02-26	Rothberg et al.	
	167	7351557		2008-04-01	Kurn	
	168	7354717		2008-04-08	Kurn	
	169	7361466		2008-04-22	Korlach et al.	
	170	7361468		2008-04-22	Liu et al.	
	171	7402386		2008-07-22	Kurn et al.	
	172	7405281		2008-07-29	Xu et al.	
	173	7414117		2008-08-19	Saito et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	174	7416844		2008-08-26	Korlach et al.	
	175	7462452		2008-12-09	Williams et al.	
	176	7462468		2008-12-09	Williams et al.	
	177	7476503		2009-01-13	Turner et al.	
	178	7476504		2009-01-13	Turner	
	179	7491498		2009-02-17	Lapidus et al.	
	180	7501245		2009-03-10	Quake et al.	
	181	7579153		2009-08-25	Brenner et al.	
	182	7704687		2010-04-27	Wang et al.	
	183	7741463		2010-06-22	Gormley et al.	
	184	7771934		2010-08-10	Kurn	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	185	7771946		2010-08-10	Kurn	
	186	7803550		2010-09-28	Makarov et al.	
	187	7846666		2010-12-07	Kurn	
	188	7846733		2010-12-07	Kurn	
	189	7867703		2011-01-11	Sampson et al.	
	190	7939258		2011-05-10	Kurn et al.	
	191	7948015		2011-05-24	Rothberg et al.	
	192	7985565		2011-07-26	Mayer et al.	
	193	8017335		2011-09-13	Smith	
	194	8034568		2011-10-11	Kurn et al.	
	195	8053192		2011-11-08	Bignell et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	196	8071311		2011-12-06	Kurn	
	197	8143001		2012-03-27	Kurn et al.	
	198	8209130		2012-06-26	Kennedy et al.	
	199	8334116		2012-12-18	Kurn	
	200	8465950		2013-06-18	Kurn et al.	
	201	8492095		2013-07-23	Kurn	
	202	8512956		2013-08-20	Kurn	
	203	8551709		2013-10-08	Kurn et al.	
	204	8852867		2014-10-07	Kurn et al.	
	205	8999677		2015-04-07	Soldatov et al.	
	206	9175325		2015-11-03	Kurn et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	207	9175336		2015-11-03	Soldatov et al.	
	208	9181582		2015-11-10	Kurn	
	209	9206418		2015-12-08	Armour	
	210	9248076		2016-02-02	Sullivan et al.	
	211	9546399		2017-01-17	Amorese et al.	
	212	9702004		2017-07-11	Van Eijk et al.	
	213	9745627		2017-08-29	van Eijk et al.	
	214	9896721		2018-02-20	Van Eijk et al.	

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	20010000077	A1	2001-03-29	Engelhardt et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	2	20010031739	A1	2001-10-18	Dare	
	3	20010034048	A1	2001-10-25	Kurn	
	4	20010041334	A1	2001-11-15	Rashtchian et al.	
	5	20020028447	A1	2002-03-07	Li et al.	
	6	20020058270	A1	2002-05-16	Kurn	
	7	20020115088	A1	2002-08-22	Kurn	
	8	20020150919	A1	2002-10-17	Weismann et al.	
	9	20020155451	A1	2002-10-24	Makrigiorgos	
	10	20020164628	A1	2002-11-07	Kurn	
	11	20020164634	A1	2002-11-07	Patil et al.	
	12	20020197639	A1	2002-12-26	Shia et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	13	20030017591	A1	2003-01-23	Kurn	
	14	20030022207	A1	2003-01-30	Balasubramanian et al.	
	15	20030082543	A1	2003-05-01	Su et al.	
	16	20030087251	A1	2003-05-08	Kurn	
	17	20030119150	A1	2003-06-26	Ankenbauer et al.	
	18	20030143555	A1	2003-07-31	Bourget et al.	
	19	20030175780	A1	2003-09-18	Jones	
	20	20030180779	A1	2003-09-25	Lofton-Day et al.	
	21	20030186234	A1	2003-10-02	Kurn	
	22	20030207279	A1	2003-11-06	Crothers et al.	
	23	20030211616	A1	2003-11-13	Leong	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	24	20030215926	A1	2003-11-20	Kurn et al.	
	25	20030224439	A1	2003-12-04	Lafferty et al.	
	26	20030232348	A1	2003-12-18	Jones et al.	
	27	20040002371	A1	2004-01-01	Paquin et al.	
	28	20040005614	A1	2004-01-08	Kurn et al.	
	29	20040023271	A1	2004-02-05	Kurn et al.	
	30	20040115815	A1	2004-06-17	Li et al.	
	31	20040137456	A1	2004-07-15	Yokota et al.	
	32	20040161742	A1	2004-08-19	Dean et al.	
	33	20040203019	A1	2004-10-14	Kurn	
	34	20040203025	A1	2004-10-14	Kurn	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	35	20040248153	A1	2004-12-09	Dear et al.	
	36	20050003441	A1	2005-01-06	Kurn	
	37	20050014192	A1	2005-01-20	Kurn	
	38	20050019793	A1	2005-01-27	Kurn et al.	
	39	20050059048	A1	2005-03-17	Gunderson et al.	
	40	20050064456	A1	2005-03-24	Kurn	
	41	20050123956	A1	2005-06-09	Blume et al.	
	42	20050136417	A1	2005-06-23	Cole et al.	
	43	20050142577	A1	2005-06-30	Jones et al.	
	44	20050191656	A1	2005-09-01	Drmanac et al.	
	45	20050191682	A1	2005-09-01	Barone et al.	

IDT Exhibit 1019.0324

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	46	20050208538	A1	2005-09-22	Kurn et al.	
	47	20060008824	A1	2006-01-12	Ronaghi et al.	
	48	20060014182	A1	2006-01-19	Kurn	
	49	20060024678	A1	2006-02-02	Buzby	
	50	20060024711	A1	2006-02-02	Lapidus et al.	
	51	20060035274	A1	2006-02-16	Dong	
	52	20060046251	A1	2006-03-02	Sampson et al.	
	53	20060051789	A1	2006-03-09	Kazakov et al.	
	54	20060068415	A1	2006-03-30	Jones et al.	
	55	20060134633	A1	2006-06-22	Chen et al.	
	56	20060216724	A1	2006-09-28	Christians et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	57	20060263789	A1	2006-11-23	Kincaid	
	58	20060281082	A1	2006-12-14	Zhu	
	59	20060286566	A1	2006-12-21	Lapidus et al.	
	60	20060292597	A1	2006-12-28	Shapero et al.	
	61	20070031857	A1	2007-02-08	Makarov et al.	
	62	20070134128	A1	2007-06-14	Korlach	
	63	20070141604	A1	2007-06-21	Gormley et al.	
	64	20070224607	A1	2007-09-27	Morgan et al.	
	65	20070224613	A1	2007-09-27	Strathmann	
	66	20070231823	A1	2007-10-04	McKernan et al.	
	67	20070238122	A1	2007-10-11	Allbritton et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	68	20070263045	A1	2007-11-15	Okazawa	
	69	20080038727	A1	2008-02-14	Spier	
	70	20080087826	A1	2008-04-17	Harris et al.	
	71	20080103058	A1	2008-05-01	Siddiqi	
	72	20080131937	A1	2008-06-05	Schroeder	
	73	20080160580	A1	2008-07-03	Adessi et al.	
	74	20080176311	A1	2008-07-24	Kurn	
	75	20080182300	A1	2008-07-31	Kurn	
	76	20080194413	A1	2008-08-14	Albert	
	77	20080194416	A1	2008-08-14	Chen	
	78	20080206764	A1	2008-08-28	Williams et al.	

IDT Exhibit 1019.0327

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	79	20080213770	A1	2008-09-04	Williams et al.	
	80	20080217246	A1	2008-09-11	Benn et al.	
	81	20080241831	A1	2008-10-02	Fan et al.	
	82	20080242560	A1	2008-10-02	GUNDERSON et al.	
	83	20080286795	A1	2008-11-20	Kawashima et al.	
	84	20090011959	A1	2009-01-08	Costa et al.	
	85	20090024331	A1	2009-01-22	Tomaney et al.	
	86	20090026082	A1	2009-01-29	Rothberg 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	2444926	CA	A1	2002-11-14	Centre Nat Rech Scient		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	2	101565746	CN	A	2009-10-28	Univ Southeast		☒
	3	105890722	CN	A	2016-08-24	Univ Changzhou		☒
	4	1661102	CN	A	2005-08-31	Univ Shanghai Jiaotong		☒
	5	0329822	EP	B1	1994-06-08	Cangene Corp		☐
	6	0365627	EP	B1	1993-12-22	Univ Iowa Res Found		☐
	7	0667393	EP	A2	1995-08-16	Enzo Diagnostics, Inc		☐
	8	0843735	EP	B1	2002-07-17	Forfas Trading As Biores Irela		☐
	9	1071811	EP	B1	2002-03-06	Entpr Ie Trd As Bioresearch Ie		☐
	10	1929039	EP	B2	2013-11-20	Keygene NV		☐
	11	2272976	EP	A1	2011-01-12	Max Planck Gesellschaft		☐
	12	2322612	EP	A1	2011-05-18	Hitachi High Tech Corp		☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	13	2451973	EP	A1	2012-05-16	Max Planck Gesellschaft		☐
	14	2511381	EP	A1	2012-10-17	Monsanto Technology LLC		☐
	15	00/018957	WO	A1	2000-04-06	Applied Research Systems		☐
	16	00/08208	WO	A2	2000-02-17	Medical Res Council		☐
	17	00/39345	WO	A1	2000-07-06	Dana-farber Cancer Institute, Inc		☐
	18	00/52191	WO	A1	2000-09-08	Life Technologies, Inc		☐
	19	00/70039	WO	A1	2000-11-23	Univ Mcgill		☐
	20	01/20035	WO	A2	2001-03-22	Nugen Technologies, Inc		☐
	21	01/23613	WO	A1	2001-04-05	Isis Pharmaceuticals, Inc		☐
	22	01/46464	WO	A1	2001-06-28	Lion Bioscience AG		☐
	23	01/57248	WO	A2	2001-08-09	Solexa Ltd		☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	24	01/64952	WO	A2	2001-09-07	Molecular Staging, Inc		☐
	25	02/00938	WO	A2	2002-01-03	Nugen Tgechnologies, Inc		☐
	26	02/060318	WO	A2	2002-08-08	Univ Ohio State Res Found		☐
	27	02/072772	WO	A2	2002-09-19	Nugen Technologies, Inc		☐
	28	02/072773	WO	A2	2002-09-19	Nugen Technologies, Inc		☐
	29	02/081753	WO	A1	2002-10-17	Advanced Res & Tech Inst		☐
	30	02/090584	WO	A2	2002-11-14	Bio Merieux		☐
	31	02/28876	WO	A2	2002-04-11	Riken		☐
	32	02/29117	WO	A2	2002-04-11	Nugen Technologies, Inc		☐
	33	02/36821	WO	A2	2002-05-10	University College Cork - National University Of I		☐
	34	02/48402	WO	A2	2002-06-20	Nugen Technologies, Inc		☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	35	03/004690	WO	A2	2003-01-16	454 Corp		☐
	36	03/027259	WO	A2	2003-04-03	Epigenx Pharmaceuticals, Inc		☐
	37	03/078645	WO	A2	2003-09-25	Nugen Technologies, Inc		☐
	38	03/083435	WO	A2	2003-10-09	Nugen Technologies, Inc		☐
	39	03/106642	WO	A2	2003-12-24	Affymetrix, Inc		☐
	40	04/011665	WO	A2	2004-02-05	Nugen Technologies, Inc		☐
	41	2000/09756	WO	A1	2000-02-24	Perkin Elmer Corp		☐
	42	2000/55364	WO	A2	2000-09-21	Amersham Pharm Biotech UK Ltd		☐
	43	2003/002736	WO	A2	2003-01-09	Roche Diagnostics GmbH		☐
	44	2003/012118	WO	A1	2003-02-13	Affymetrix, Inc		☐
	45	2004/070007	WO	A2	2004-08-19	454 Corp		☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	46	2004/092418	WO	A2	2004-10-28	Nugen Technologies, Inc		☐
	47	2005/003375	WO	A2	2005-01-13	454 Corp		☐
	48	2005/038427	WO	A2	2005-04-28	Medical Gene Ct Ltd		☐
	49	2005/065321	WO	A2	2005-07-21	Nugen Technologies, Inc		☐
	50	2006/081222	WO	A2	2006-08-03	Compass Genetics, LLC		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

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	Amos, 2000, DNA pooling in mutation detection with reference to sequence analysis, Am J Hum Genet 66:1689-1692	
	2	Bellos, 2014, cnvCapSeq: detecting copy number variation in long-range targeted resequencing data, Nucleic Acids Res 42(20):e158	
	3	Benson, 2013, Genbank, Nucl Acids Res 41:D36-D42	
	4	Blomquist, 2013, Targeted RNA-Sequencing with Competitive Multiplex-PCR Amplicon Libraries, PLOS ONE 8(11): e79120	

IDT Exhibit 1019.0333

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	5	Bodi, 2013, Comparison of commercially available target enrichment methods for next-generation sequencing, J Biomolecular Tech 24:73-86
	6	Church, 1988, Multiplexed DNA sequencing, Science 240:185-188
	7	Colbert, 2001, High-throughput screening for induced point mutations, Plant Physiol 126:480-484
	8	Collard, 2005, An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvements: the basic concepts Euphytica 142:169-196
	9	Eminaga, 2013, Quantification of microRNA Expression with Next-Generation Sequencing, Unit 4.17 in Current Protocols in Molecular Biology, Wiley, New York, NY (14 pages)
	10	Fakhrai-Rad, 2002, Pyroseqeuncing: An accurate detection platform for single nucleotide polymorphisms, Human Mutation 19:479-485
	11	Grothues, 1993, PCR amplification of megabase DNA with tagged random primers (T-PCR), Nucl Acids Res 21:1321-1322
	12	Hajibabaei, 2005, Critical factors for assembling a high volume of DNA barcodes, Phil Trans R Soc B 360:1959-1967
	13	Illumina, 2011, TruSeq RNA and DNA Sample Preparation Kits v2, 1-15 Illumina, dated 27 April 27, 2011 (4 pages)
	14	International Search Report and Written Opinion mailed January 6, 2016, for PCT/US15/44065, filed August 6, 2015 (21 pages)
	15	International Search Report and Written Opinion mailed July 10, 2017, for Application No. PCT/US17/27060, filed April 11, 2017 (9 pages)

IDT Exhibit 1019.0334

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	16	International Search Report and Written Opinion mailed Mar. 5, 2015, in international patent application PCT/US2014/065530, filed November 13, 2014 (12 pages)
	17	Ion Total RNA-Seq Kit v2, User Guide, 2012, Life Technologies (82 pages)
	18	Jiang, 2015, CODEX: a normalization and copy number variation detection method for whole exome sequencing, Nucleic Acids Res 43(6):e39
	19	Krumm, 2012, Copy number variation detection and genotyping from exome sequence data, Genome Res 22 (8):1525-1532
	20	Lai, 2004, Characterization of the maize endosperm transcriptome and its comparison to the rice genome, Genome Res14:1932-1937
	21	Langmead, 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol 10:R25
	22	Li, 2012, CONTRA: copy number analysis for targeted resequencing, Bioinformatics 28(10):1307-1313
	23	Lindstrom, 2004, Pyrosequencing for detection of Lamivudine-resistant Hepatitis B virus, J Clin Microb 42 (10):4788-4795
	24	Liu, 2008, Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples, BMC Genomics 9(509):1-17
	25	Ma, 2015, Quantitative Analysis of Copy Number Variants Based on Real-Time LightCycler PCR, Curr Protoc Hum Genet 80:7.21.1-7.23.8
	26	Machine translation generated on March 7, 2018, of CN 105890722 by website of European Patent Office (4 pages)

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	27	Margulies, 2005, Genorne sequencing in open microfabricated high density picoliter reactors, Nature 437 (7057):376-380
	28	McCloskey, 2007, Encoding PCR products with batch-stamps and barcodes, Biochem Genet 45:761-767
	29	Miner, 2004, Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR, Nucl Acids Res 32 (17):e135
	30	Myers, 2013, Protocol for Creating Multiplexed miRNA Libraries for Use in Illumina Sequencing, Myers lab microRNA-seq Protocol, Hudson Alpha Institute for Biotechnology web site, dated May 2, 2013, (15 pages)
	31	NuGEN, 2014, User Guide Ovation Target Enrichment System, NuGEN Technologies Inc., San Carlos, CA (45 pages)
	32	Plagnol, 2012, A robust model for read count data in exome sequencing experiments and implications for copy number variant calling, Bioinformatics 28(21):2747-2754
	33	Qiu, 2003, DNA sequence-based "bar-codes" for tracking the origins of expressed sequence tags from a maize cDNA library constructed using multiple mRNA sources, Plant Physiol 133:475-481
	34	Querfurth, 2012, Creation and application of immortalized bait libraries for targeted enrichment and next-generation sequencing, Biotechniques 52(6):375-380
	35	Ronaghi, 2001, Pyrosequencing sheds light on DNA sequencing, Genome Res 11:3-11
	36	Sathirapongsasuti, 2011, Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV, Bioinformatics 27(19):2648-2654
	37	Schiemer, 2011, Illumina TruSeq Adapters Demystified,TUFTS UNIVERSITY CORE FACILITY XP055357867 (5 pages)

IDT Exhibit 1019.0336

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	38	Shapero, 2001, SNP Genotyping by multiplexed solid-phase amplification and fluorescent minisequencing, Genome Res 11:1926-1934
	39	Shendure, 2005, Accurate multiplex polony sequencing of an evolved bacterial genome, Science 309:1728
	40	Soni, 2007, Progress toward ultrafast DNA sequencing using solid-state nanopores, Clin Chem 53(11):1996-2001
	41	Sood, 2006, Methods for reverse genetic screening in zebrafish by resequencing and TILLING, Methods 39:220-227
	42	Staroscik, 2004, Calculator for determining the number of copies of a template, URI Genomics, webpage archive dated April 6, 2017 (1 page), Retreived from the internet on March 7, 2018, from <https://web.archive.org/web/20170406174850/http://cels.uri.edu/gsc/cndna.html>
	43	Stratagene, 1998, Gene characterization kits, Stratagene Catalog, p. 39 (2 pages)
	44	Supplementary European search report and opinion mailed January 30, 2018, for European patent application No. 15830393.3 (6 pages)
	45	Till, 2003, Large-scale discovery of induced point mutations with high-throughput TILLING, Genome Res 13:524-530
	46	Trapnell, 2010, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat Biotech 28:511-515
	47	Trapnell, 2013, Differential analysis of gene regulation at transcript resolution with RNA-seq, Nat Biotech 31:46-53
	48	Unemo, 2004, Molecular typing of Neisseria gonorrhoeae isolates by pyrosequencing of highly polymorphic segments of the porB gene, J Clin Microb 42(7):2926-2934

IDT Exhibit 1019.0337

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	49	Vigal, 2002, A review on SNP and other types of molecular markers and their use in animal genetics, Genet Sel Evol 34:275-305
	50	Walker, 1992, Strand displacement amplification–an isothermal, in vitro DNA amplification technique, Nucl Acids Res 20(7):1691-1696

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2020-07-22
Name/Print	Thomas C. Meyers	Registration Number	36,989

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.**

IDT Exhibit 1019.0339

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.

IDT Exhibit 1019.0340

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
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516

CONFIRMATION NO. 3249

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

PUBLICATION NOTICE

OC000000118844208

Title:COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

Publication No.US-2020-0239953-A1
Publication Date:07/30/2020

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

IDT Exhibit 1019.0341

Electronic Acknowledgement Receipt

EFS ID:	40672119
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	25-SEP-2020
Filing Date:	15-JAN-2020
Time Stamp:	14:56:21
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	NUGN-001-03US_3-73c.pdf	117852 b81aae863a20f496cc401f18a2b2ae0e72034202	no	3

Warnings:

Information:					
2	Power of Attorney	NUGN-001-03US_POA.pdf	108576 0d367b4f584b8c1a638bb6866b311f7619c69af5	no	1
Warnings:					
Information:					
		Total Files Size (in bytes):	226428		

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: NuGEN Technologies, Inc.

Application No./Patent No.: 16/743,604 Filed/Issue Date: January 15, 2020

Titled: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

NuGEN Technologies, Inc., a corporation

(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 053605, Frame 0123, 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.

/Thomas C. Meyers/	September 25, 2020
Signature	Date
Thomas C. Meyers	36,989
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.

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: 21710

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: 21710

OR

☐	Firm or Individual Name			
	Address			
	City		State	Zip
	Country			
	Telephone		Email	

Assignee Name and Address: NUGEN TECHNOLOGIES, INC.
201 INDUSTRIAL ROAD
SAN CARLOS, CA 94070

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	[signature]	Date	Sept 11, 2017
Name	Douglas A Amorese	Telephone	650 590-3615
Title	Chief Scientific Officer - NuGEN Technologies		

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.

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
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516

CONFIRMATION NO. 3249

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

POA ACCEPTANCE LETTER

OC000000120237945

Date Mailed: 09/30/2020

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 09/25/2020.

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**.

/yteferra/

IDT Exhibit 1019.0348

PALM INTRANET

Day : Wednesday

Date: 12/2/2020
Time: 12:56:54

Inventor Name Search

Enter the **first few letters** of the Inventor's Last Name.
Additionally, enter the **first few letters** of the Inventor's First name.

Last Name	First Name	
Amorese	Douglas	Search

To Go BACK Use BACK Button on Your BROWSER Tool Bar
Back to PALM | ASSIGNMENT | OASIS | Home page

IDT Exhibit 1019.0349

PALM INTRANET

Day : Wednesday

Date: 12/2/2020
Time: 12:56:54

Inventor Name Search

Enter the **first few letters** of the Inventor's Last Name.
Additionally, enter the **first few letters** of the Inventor's First name.

Last Name	First Name	
Scolnick	Jonathan	Search

To Go BACK Use BACK Button on Your BROWSER Tool Bar
Back to PALM | ASSIGNMENT | OASIS | Home page

IDT Exhibit 1019.0350

PALM INTRANET

Day : Wednesday

Date: 12/2/2020
Time: 12:56:54

Inventor Name Search

Enter the **first few letters** of the Inventor's Last Name.
Additionally, enter the **first few letters** of the Inventor's First name.

Last Name	First Name	
Schroeder	Benjamin	Search

To Go BACK Use BACK Button on Your BROWSER Tool Bar
Back to PALM | ASSIGNMENT | OASIS | Home page

IDT Exhibit 1019.0351

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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					

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	2013/190441	WO	A2	2013-12-27	Univ Hong Kong Chinese		
	2	2013/191775	WO	A2	2013-12-27	Nugen Technologies, Inc		
	3	2014/039556	WO	A1	2014-03-13	Guardant Health, Inc		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	4	2014/082032	WO	A1	2014-05-30	Univ Toledo		
	5	2014/144092	WO	A1	2014-09-18	Nugen Technologies, Inc		
	6	2014/150931	WO	A1	2014-09-25	Nugen Technologies, Inc		
	7	2015/031691	WO	A1	2015-03-05	Cellular Research, Inc		
	8	2015/073711	WO	A1	2015-05-21	Nugen Technologies, Inc		
	9	2015/131107	WO	A1	2015-09-03	Nugen Technologies, Inc		
	10	89/09284	WO	A1	1989-10-05	Univ Iowa Res Found		
	11	92/07951	WO	A1	1992-05-14	Kow, Yoke		
	12	93/18052	WO	A1	1993-09-16	Isis Pharmaceuticals, Inc		
	13	94/16090	WO	A1	1994-07-21	Molecular Tool, Inc		
	14	96/40998	WO	A1	1996-12-19	Pioneer Hi-bred International, Inc		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	15	97/12061	WO	A1	1997-04-03	Epicentre Technologies Corp		
	16	97/25416	WO	A2	1997-07-17	Nyfotek As		
	17	98/044151	WO	A1	1998-10-08	Glaxo Group Ltd		
	18	98/06736	WO	A1	1998-02-19	Life Technologies, Inc		
	19	98/38296	WO	A1	1998-09-03	Gesher Israel Advanced Biotecs		
	20	99/10540	WO	A1	1999-03-04	Lopez, Osvaldo		
	21	99/11819	WO	A1	1999-03-11	Perkin Elmer Corp		
	22	99/42618	WO	A1	1999-08-26	Dade Behring Inc		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

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		

If you wish to add additional non-patent literature document citation information please click the Add button Add

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

EXAMINER SIGNATURE			
Examiner Signature	/TERESA E STRZELECKA/	Date Considered	12/02/2020

*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. /T.E.S/

IDT Exhibit 1019.0355

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	Art Unit
	TERESA E STRZELECKA	1637

CPC - Searched*		
Symbol	Date	Examiner
C12Q1/6869 |C12Q2525/191 | C12Q2535/122 plus text search	12/02/2020	/TS/

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
Inventor name search	12/02/2020	/TS/
WEST	12/02/2020	/TS/
Google	12/02/2020	/TS/
IDS (7/22/20)	12/02/2020	/TS/

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

IDT Exhibit 1019.0356

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	Art Unit
	TERESA E STRZELECKA	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	12/03/2020								
	22	✓								
	23	✓								
	24	✓								
	25	✓								
	26	✓								
	27	✓								
	28	✓								
	29	✓								
	30	✓								
	31	✓								
	32	✓								

IDT Exhibit 1019.0357

Bibliographic Data

Application No: 16/743,604

Foreign Priority claimed: ○ Yes ◉ No

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

Verified and Acknowledged: /TERESA E STRZELECKA/ Examiner's Signature | Initials

Title: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
01/15/2020	435	1637	NUGN-001/03US 34079/516
RULE			

APPLICANTS

NUGEN TECHNOLOGIES, INC., Redwood City, CA,

INVENTORS

Douglas A. Amorese, Los Altos, CA, UNITED STATES

Jonathan Scolnick, San Francisco, CA, UNITED STATES

Benjamin G. Schroeder, San Mateo, CA, UNITED STATES

CONTINUING DATA

This application is a CON of 15406002 01/13/2017 PAT 10570448

15406002 is a DIV of 14540917 11/13/2014 PAT 9546399

14540917 has PRO of 61903826 11/13/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

02/14/2020

STATE OR COUNTRY

UNITED STATES

ADDRESS

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111
UNITED STATES

FILING FEE RECEIVED

$1,820

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	4362867		1982-12-07	Paddock	
	2	4458066		1984-07-03	Caruthers et al.	
	3	4469863		1984-09-04	Ts'o et al.	
	4	4582877		1986-04-15	Fairchok et al.	
	5	4876187		1989-10-24	Duck et al.	
	6	4925065		1990-05-15	Golias	
	7	4935357		1990-06-19	Szybalski	
	8	4942124		1990-07-17	Church	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	9	4988617		1991-01-29	Landegren et al.	
	10	4996143		1991-02-26	Heller et al.	
	11	5011769		1991-04-30	Duck et al.	
	12	5034506		1991-07-23	Summerton et al.	
	13	5035996		1991-07-30	Hartley	
	14	5043272		1991-08-27	Hartley	
	15	5082830		1992-01-21	Brakel et al.	
	16	5090591		1992-02-25	Long	
	17	5130238		1992-07-14	Malek et al.	
	18	5169766		1992-12-08	Schuster et al.	
	19	5171534		1992-12-15	Smith et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	20	5194370		1993-03-16	Berninger et al.	
	21	5216141		1993-06-01	Benner	
	22	5234809		1993-08-10	Boom et al.	
	23	5235033		1993-08-10	Summerton et al.	
	24	5242794		1993-09-07	Whiteley et al.	
	25	5384242		1995-01-24	Oakes	
	26	5386023		1995-01-31	Sanghvi et al.	
	27	5399491		1995-03-21	Kacian et al.	
	28	5409818		1995-04-25	Davey et al.	
	29	5418149		1995-05-23	Gelfand et al.	
	30	5422271		1995-06-06	Chen et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0361

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	31	5427929		1995-06-27	Richards et al.	
	32	5480784		1996-01-02	Kacian et al.	
	33	5494810		1996-02-27	Barany et al.	
	34	5508169		1996-04-16	Deugau et al.	
	35	5508178		1996-04-16	Rose et al.	
	36	5510270		1996-04-23	Fodor et al.	
	37	5525471		1996-06-11	Zeng	
	38	5545522		1996-08-13	Van Gelder et al.	
	39	5554516		1996-09-10	Kacian et al.	
	40	5554517		1996-09-10	Davey et al.	
	41	5556752		1996-09-17	Lockhart et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	42	5565340		1996-10-15	Chenchik et al.	
	43	5573913		1996-11-12	Rosemeyer et al.	
	44	5578832		1996-11-26	Trulson et al.	
	45	5589339		1996-12-31	Hampson et al.	
	46	5602240		1997-02-11	De Mesmaeker et al.	
	47	5637684		1997-06-10	Cook et al.	
	48	5641658		1997-06-24	Adams et al.	
	49	5644048		1997-07-01	Yau	
	50	5665549		1997-09-09	Pinkel et al.	
	51	5667976		1997-09-16	Van Ness et al.	
	52	5667979		1997-09-16	Berrens	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	53	5679512		1997-10-21	Laney et al.	
	54	5681726		1997-10-28	Huse et al.	
	55	5683879		1997-11-04	Laney et al.	
	56	5688648		1997-11-18	Mathies et al.	
	57	5705628		1998-01-06	Hawkins	
	58	5708154		1998-01-13	Smith et al.	
	59	5710028		1998-01-20	Eyal et al.	
	60	5712126		1998-01-27	Weissman et al.	
	61	5716785		1998-02-10	Van Gelder et al.	
	62	5726329		1998-03-10	Jones et al.	
	63	5750341		1998-05-12	Macevicz	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	64	5759822		1998-06-02	Chenchik et al.	
	65	5763178		1998-06-09	Chirikjian et al.	
	66	5789206		1998-08-04	Tavtigian et al.	
	67	5824517		1998-10-20	Cleuziat et al.	
	68	5824518		1998-10-20	Kacian et al.	
	69	5837832		1998-11-17	Chee et al.	
	70	5876976		1999-03-02	Richards et al.	
	71	5882867		1999-03-16	Ullman et al.	
	72	5888779		1999-03-30	Kacian et al.	
	73	5888819		1999-03-30	Goelet et al.	
	74	5945313		1999-08-31	Hartley et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	75	5952176		1999-09-14	McCarthy et al.	
	76	5958681		1999-09-28	Wetmur et al.	
	77	5965409		1999-10-12	Pardee et al.	
	78	5969119		1999-10-19	Macevicz	
	79	5972618		1999-10-26	Bloch	
	80	6004744		1999-12-21	Goelet et al.	
	81	6004745		1999-12-21	Arnold, Jr. et al.	
	82	6027889		2000-02-22	Barany et al.	
	83	6027923		2000-02-22	Wallace	
	84	6030774		2000-02-29	Laney et al.	
	85	6037152		2000-03-14	Richards et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0366

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	86	6056661		2000-05-02	Schmidt	
	87	6077674		2000-06-20	Schleifer et al.	
	88	6087103		2000-07-11	Burmer	
	89	6090553		2000-07-18	Matson	
	90	6090591		2000-07-18	Burg et al.	
	91	6107023		2000-08-22	Reyes et al.	
	92	6110709		2000-08-29	Ausubel et al.	
	93	6150112		2000-11-21	Weissman et al.	
	94	6159685		2000-12-12	Pinkel et al.	
	95	6160105		2000-12-12	Cunningham et al.	
	96	6169194		2001-01-02	Thompson et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	97	6172208		2001-01-09	Cook	
	98	6174680		2001-01-16	Makrigiorgos	
	99	6180338		2001-01-30	Adams	
	100	6190865		2001-02-20	Jendrisak et al.	
	101	6194211		2001-02-27	Richards et al.	
	102	6197501		2001-03-06	Cremer et al.	
	103	6197557		2001-03-06	Makarov et al.	
	104	6210891		2001-04-03	Nyren et al.	
	105	6225109		2001-05-01	Juncosa et al.	
	106	6225451		2001-05-01	Ballinger et al.	
	107	6232104		2001-05-15	Lishanski et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	108	6251639		2001-06-26	Kurn	
	109	6262490		2001-07-17	Hsu et al.	
	110	6270961		2001-08-07	Drmanac	
	111	6280935		2001-08-28	Macevicz	
	112	6287766		2001-09-11	Nolan et al.	
	113	6287825		2001-09-11	Weissman et al.	
	114	6291170		2001-09-18	Van Gelder et al.	
	115	6306365		2001-10-23	Ruoslahti et al.	
	116	6306597		2001-10-23	Macevicz	
	117	6309843		2001-10-30	Timms	
	118	6326142		2001-12-04	Royer	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	119	6335167		2002-01-01	Pinkel et al.	
	120	6339147		2002-01-15	Lukhtanov et al.	
	121	6440705		2002-08-27	Stanton, Jr. et al.	
	122	6449562		2002-09-10	Chandler et al.	
	123	6582938		2003-06-24	Su et al.	
	124	6670461		2003-12-30	Wengel et al.	
	125	6686156		2004-02-03	Kurn	
	126	6692918		2004-02-17	Kurn	
	127	6770748		2004-08-03	Imanishi et al.	
	128	6777180		2004-08-17	Fisher et al.	
	129	6794499		2004-09-21	Wengel et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	130	6815164		2004-11-09	Kurn	
	131	6815167		2004-11-09	Crothers et al.	
	132	6825011		2004-11-30	Romantchikov	
	133	6833246		2004-12-21	Balasubramanian	
	134	6849404		2005-02-01	Park et al.	
	135	6858413		2005-02-22	Kurn	
	136	6913884		2005-07-05	Stuelpnagel et al.	
	137	6917726		2005-07-12	Levene et al.	
	138	6924104		2005-08-02	Weissman et al.	
	139	6946251		2005-09-20	Kurn	
	140	7001724		2006-02-21	Greenfield	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	141	7033764		2006-04-25	Korlach et al.	
	142	7048481		2006-05-23	Sugata et al.	
	143	7052847		2006-05-30	Korlach et al.	
	144	7056676		2006-06-06	Korlach et al.	
	145	7056716		2006-06-06	Potter et al.	
	146	7060441		2006-06-13	Bourget et al.	
	147	7094536		2006-08-22	Kurn	
	148	7115400		2006-10-03	Adessi et al.	
	149	7170050		2007-01-30	Turner et al.	
	150	7175982		2007-02-13	McCarthy et al.	
	151	7176025		2007-02-13	Kurn et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	152	7189512		2007-03-13	Porat et al.	
	153	7211390		2007-05-01	Rothberg et al.	
	154	7232656		2007-06-19	Balasubramanian et al.	
	155	7244559		2007-07-17	Rothberg et al.	
	156	7244567		2007-07-17	Chen et al.	
	157	7264929		2007-09-04	Rothberg et al.	
	158	7273730		2007-09-25	Du Breuil Lastrucci	
	159	7276720		2007-10-02	Ulmer	
	160	7294461		2007-11-13	Kurn	
	161	7300755		2007-11-27	Petersdorf et al.	
	162	7302146		2007-11-27	Turner et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	163	7313308		2007-12-25	Turner et al.	
	164	7315019		2008-01-01	Turner et al.	
	165	7323305		2008-01-29	Leamon et al.	
	166	7335762		2008-02-26	Rothberg et al.	
	167	7351557		2008-04-01	Kurn	
	168	7354717		2008-04-08	Kurn	
	169	7361466		2008-04-22	Korlach et al.	
	170	7361468		2008-04-22	Liu et al.	
	171	7402386		2008-07-22	Kurn et al.	
	172	7405281		2008-07-29	Xu et al.	
	173	7414117		2008-08-19	Saito et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	174	7416844		2008-08-26	Korlach et al.	
	175	7462452		2008-12-09	Williams et al.	
	176	7462468		2008-12-09	Williams et al.	
	177	7476503		2009-01-13	Turner et al.	
	178	7476504		2009-01-13	Turner	
	179	7491498		2009-02-17	Lapidus et al.	
	180	7501245		2009-03-10	Quake et al.	
	181	7579153		2009-08-25	Brenner et al.	
	182	7704687		2010-04-27	Wang et al.	
	183	7741463		2010-06-22	Gormley et al.	
	184	7771934		2010-08-10	Kurn	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	185	7771946		2010-08-10	Kurn	
	186	7803550		2010-09-28	Makarov et al.	
	187	7846666		2010-12-07	Kurn	
	188	7846733		2010-12-07	Kurn	
	189	7867703		2011-01-11	Sampson et al.	
	190	7939258		2011-05-10	Kurn et al.	
	191	7948015		2011-05-24	Rothberg et al.	
	192	7985565		2011-07-26	Mayer et al.	
	193	8017335		2011-09-13	Smith	
	194	8034568		2011-10-11	Kurn et al.	
	195	8053192		2011-11-08	Bignell et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	196	8071311		2011-12-06	Kurn	
	197	8143001		2012-03-27	Kurn et al.	
	198	8209130		2012-06-26	Kennedy et al.	
	199	8334116		2012-12-18	Kurn	
	200	8465950		2013-06-18	Kurn et al.	
	201	8492095		2013-07-23	Kurn	
	202	8512956		2013-08-20	Kurn	
	203	8551709		2013-10-08	Kurn et al.	
	204	8852867		2014-10-07	Kurn et al.	
	205	8999677		2015-04-07	Soldatov et al.	
	206	9175325		2015-11-03	Kurn et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	207	9175336		2015-11-03	Soldatov et al.	
	208	9181582		2015-11-10	Kurn	
	209	9206418		2015-12-08	Armour	
	210	9248076		2016-02-02	Sullivan et al.	
	211	9546399		2017-01-17	Amorese et al.	
	212	9702004		2017-07-11	Van Eijk et al.	
	213	9745627		2017-08-29	van Eijk et al.	
	214	9896721		2018-02-20	Van Eijk et al.	

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	20010000077	A1	2001-03-29	Engelhardt et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	2	20010031739	A1	2001-10-18	Dare	
	3	20010034048	A1	2001-10-25	Kurn	
	4	20010041334	A1	2001-11-15	Rashtchian et al.	
	5	20020028447	A1	2002-03-07	Li et al.	
	6	20020058270	A1	2002-05-16	Kurn	
	7	20020115088	A1	2002-08-22	Kurn	
	8	20020150919	A1	2002-10-17	Weismann et al.	
	9	20020155451	A1	2002-10-24	Makrigiorgos	
	10	20020164628	A1	2002-11-07	Kurn	
	11	20020164634	A1	2002-11-07	Patil et al.	
	12	20020197639	A1	2002-12-26	Shia et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0379

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	13	20030017591	A1	2003-01-23	Kurn	
	14	20030022207	A1	2003-01-30	Balasubramanian et al.	
	15	20030082543	A1	2003-05-01	Su et al.	
	16	20030087251	A1	2003-05-08	Kurn	
	17	20030119150	A1	2003-06-26	Ankenbauer et al.	
	18	20030143555	A1	2003-07-31	Bourget et al.	
	19	20030175780	A1	2003-09-18	Jones	
	20	20030180779	A1	2003-09-25	Lofton-Day et al.	
	21	20030186234	A1	2003-10-02	Kurn	
	22	20030207279	A1	2003-11-06	Crothers et al.	
	23	20030211616	A1	2003-11-13	Leong	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0380

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	24	20030215926	A1	2003-11-20	Kurn et al.	
	25	20030224439	A1	2003-12-04	Lafferty et al.	
	26	20030232348	A1	2003-12-18	Jones et al.	
	27	20040002371	A1	2004-01-01	Paquin et al.	
	28	20040005614	A1	2004-01-08	Kurn et al.	
	29	20040023271	A1	2004-02-05	Kurn et al.	
	30	20040115815	A1	2004-06-17	Li et al.	
	31	20040137456	A1	2004-07-15	Yokota et al.	
	32	20040161742	A1	2004-08-19	Dean et al.	
	33	20040203019	A1	2004-10-14	Kurn	
	34	20040203025	A1	2004-10-14	Kurn	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	35	20040248153	A1	2004-12-09	Dear et al.	
	36	20050003441	A1	2005-01-06	Kurn	
	37	20050014192	A1	2005-01-20	Kurn	
	38	20050019793	A1	2005-01-27	Kurn et al.	
	39	20050059048	A1	2005-03-17	Gunderson et al.	
	40	20050064456	A1	2005-03-24	Kurn	
	41	20050123956	A1	2005-06-09	Blume et al.	
	42	20050136417	A1	2005-06-23	Cole et al.	
	43	20050142577	A1	2005-06-30	Jones et al.	
	44	20050191656	A1	2005-09-01	Drmanac et al.	
	45	20050191682	A1	2005-09-01	Barone et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	46	20050208538	A1	2005-09-22	Kurn et al.	
	47	20060008824	A1	2006-01-12	Ronaghi et al.	
	48	20060014182	A1	2006-01-19	Kurn	
	49	20060024678	A1	2006-02-02	Buzby	
	50	20060024711	A1	2006-02-02	Lapidus et al.	
	51	20060035274	A1	2006-02-16	Dong	
	52	20060046251	A1	2006-03-02	Sampson et al.	
	53	20060051789	A1	2006-03-09	Kazakov et al.	
	54	20060068415	A1	2006-03-30	Jones et al.	
	55	20060134633	A1	2006-06-22	Chen et al.	
	56	20060216724	A1	2006-09-28	Christians et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	57	20060263789	A1	2006-11-23	Kincaid	
	58	20060281082	A1	2006-12-14	Zhu	
	59	20060286566	A1	2006-12-21	Lapidus et al.	
	60	20060292597	A1	2006-12-28	Shapero et al.	
	61	20070031857	A1	2007-02-08	Makarov et al.	
	62	20070134128	A1	2007-06-14	Korlach	
	63	20070141604	A1	2007-06-21	Gormley et al.	
	64	20070224607	A1	2007-09-27	Morgan et al.	
	65	20070224613	A1	2007-09-27	Strathmann	
	66	20070231823	A1	2007-10-04	McKernan et al.	
	67	20070238122	A1	2007-10-11	Allbritton et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	68	20070263045	A1	2007-11-15	Okazawa	
	69	20080038727	A1	2008-02-14	Spier	
	70	20080087826	A1	2008-04-17	Harris et al.	
	71	20080103058	A1	2008-05-01	Siddiqi	
	72	20080131937	A1	2008-06-05	Schroeder	
	73	20080160580	A1	2008-07-03	Adessi et al.	
	74	20080176311	A1	2008-07-24	Kurn	
	75	20080182300	A1	2008-07-31	Kurn	
	76	20080194413	A1	2008-08-14	Albert	
	77	20080194416	A1	2008-08-14	Chen	
	78	20080206764	A1	2008-08-28	Williams et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	79	20080213770	A1	2008-09-04	Williams et al.	
	80	20080217246	A1	2008-09-11	Benn et al.	
	81	20080241831	A1	2008-10-02	Fan et al.	
	82	20080242560	A1	2008-10-02	GUNDERSON et al.	
	83	20080286795	A1	2008-11-20	Kawashima et al.	
	84	20090011959	A1	2009-01-08	Costa et al.	
	85	20090024331	A1	2009-01-22	Tomaney et al.	
	86	20090026082	A1	2009-01-29	Rothberg 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	2444926	CA	A1	2002-11-14	Centre Nat Rech Scient		

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	2	101565746	CN	A	2009-10-28	Univ Southeast		×
	3	105890722	CN	A	2016-08-24	Univ Changzhou		☒
	4	1661102	CN	A	2005-08-31	Univ Shanghai Jiaotong		☒
	5	0329822	EP	B1	1994-06-08	Cangene Corp		☐
	6	0365627	EP	B1	1993-12-22	Univ Iowa Res Found		☐
	7	0667393	EP	A2	1995-08-16	Enzo Diagnostics, Inc		☐
	8	0843735	EP	B1	2002-07-17	Forfas Trading As Biores Irela		☐
	9	1071811	EP	B1	2002-03-06	Entpr Ie Trd As Bioresearch Ie		☐
	10	1929039	EP	B2	2013-11-20	Keygene NV		☐
	11	2272976	EP	A1	2011-01-12	Max Planck Gesellschaft		☐
	12	2322612	EP	A1	2011-05-18	Hitachi High Tech Corp		☐

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	13	2451973	EP	A1	2012-05-16	Max Planck Gesellschaft		☐
	14	2511381	EP	A1	2012-10-17	Monsanto Technology LLC		☐
	15	00/018957	WO	A1	2000-04-06	Applied Research Systems		☐
	16	00/08208	WO	A2	2000-02-17	Medical Res Council		☐
	17	00/39345	WO	A1	2000-07-06	Dana-farber Cancer Institute, Inc		☐
	18	00/52191	WO	A1	2000-09-08	Life Technologies, Inc		☐
	19	00/70039	WO	A1	2000-11-23	Univ Mcgill		☐
	20	01/20035	WO	A2	2001-03-22	Nugen Technologies, Inc		☐
	21	01/23613	WO	A1	2001-04-05	Isis Pharmaceuticals, Inc		☐
	22	01/46464	WO	A1	2001-06-28	Lion Bioscience AG		☐
	23	01/57248	WO	A2	2001-08-09	Solexa Ltd		☐

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	24	01/64952	WO	A2	2001-09-07	Molecular Staging, Inc		☐
	25	02/00938	WO	A2	2002-01-03	Nugen Tgechnologies, Inc		☐
	26	02/060318	WO	A2	2002-08-08	Univ Ohio State Res Found		☐
	27	02/072772	WO	A2	2002-09-19	Nugen Technologies, Inc		☐
	28	02/072773	WO	A2	2002-09-19	Nugen Technologies, Inc		☐
	29	02/081753	WO	A1	2002-10-17	Advanced Res & Tech Inst		☐
	30	02/090584	WO	A2	2002-11-14	Bio Merieux		☐
	31	02/28876	WO	A2	2002-04-11	Riken		☐
	32	02/29117	WO	A2	2002-04-11	Nugen Technologies, Inc		☐
	33	02/36821	WO	A2	2002-05-10	University College Cork - National University Of I		☐
	34	02/48402	WO	A2	2002-06-20	Nugen Technologies, Inc		☐

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	35	03/004690	WO	A2	2003-01-16	454 Corp		☐
	36	03/027259	WO	A2	2003-04-03	Epigenx Pharmaceuticals, Inc		☐
	37	03/078645	WO	A2	2003-09-25	Nugen Technologies, Inc		☐
	38	03/083435	WO	A2	2003-10-09	Nugen Technologies, Inc		☐
	39	03/106642	WO	A2	2003-12-24	Affymetrix, Inc		☐
	40	04/011665	WO	A2	2004-02-05	Nugen Technologies, Inc		☐
	41	2000/09756	WO	A1	2000-02-24	Perkin Elmer Corp		☐
	42	2000/55364	WO	A2	2000-09-21	Amersham Pharm Biotech UK Ltd		☐
	43	2003/002736	WO	A2	2003-01-09	Roche Diagnostics GmbH		☐
	44	2003/012118	WO	A1	2003-02-13	Affymetrix, Inc		☐
	45	2004/070007	WO	A2	2004-08-19	454 Corp		☐

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	46	2004/092418	WO	A2	2004-10-28	Nugen Technologies, Inc		☐
	47	2005/003375	WO	A2	2005-01-13	454 Corp		☐
	48	2005/038427	WO	A2	2005-04-28	Medical Gene Ct Ltd		☐
	49	2005/065321	WO	A2	2005-07-21	Nugen Technologies, Inc		☐
	50	2006/081222	WO	A2	2006-08-03	Compass Genetics, LLC		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

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.	T5
	1	Amos, 2000, DNA pooling in mutation detection with reference to sequence analysis, Am J Hum Genet 66:1689-1692	
	2	Bellos, 2014, cnvCapSeq: detecting copy number variation in long-range targeted resequencing data, Nucleic Acids Res 42(20):e158	
	3	Benson, 2013, Genbank, Nucl Acids Res 41:D36-D42	
	4	Blomquist, 2013, Targeted RNA-Sequencing with Competitive Multiplex-PCR Amplicon Libraries, PLOS ONE 8(11): e79120	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0391

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	5	Bodi, 2013, Comparison of commercially available target enrichment methods for next-generation sequencing, J Biomolecular Tech 24:73-86
	6	Church, 1988, Multiplexed DNA sequencing, Science 240:185-188
	7	Colbert, 2001, High-throughput screening for induced point mutations, Plant Physiol 126:480-484
	8	Collard, 2005, An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvements: the basic concepts Euphytica 142:169-196
	9	Eminaga, 2013, Quantification of microRNA Expression with Next-Generation Sequencing, Unit 4.17 in Current Protocols in Molecular Biology, Wiley, New York, NY (14 pages)
	10	Fakhrai-Rad, 2002, Pyroseqeuncing: An accurate detection platform for single nucleotide polymorphisms, Human Mutation 19:479-485
	11	Grothues, 1993, PCR amplification of megabase DNA with tagged random primers (T-PCR), Nucl Acids Res 21:1321-1322
	12	Hajibabaei, 2005, Critical factors for assembling a high volume of DNA barcodes, Phil Trans R Soc B 360:1959-1967
	13	Illumina, 2011, TruSeq RNA and DNA Sample Preparation Kits v2, 1-15 Illumina, dated 27 April 27, 2011 (4 pages)
	14	International Search Report and Written Opinion mailed January 6, 2016, for PCT/US15/44065, filed August 6, 2015 (21 pages)
	15	International Search Report and Written Opinion mailed July 10, 2017, for Application No. PCT/US17/27060, filed April 11, 2017 (9 pages)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	16	International Search Report and Written Opinion mailed Mar. 5, 2015, in international patent application PCT/US2014/065530, filed November 13, 2014 (12 pages)
	17	Ion Total RNA-Seq Kit v2, User Guide, 2012, Life Technologies (82 pages)
	18	Jiang, 2015, CODEX: a normalization and copy number variation detection method for whole exome sequencing, Nucleic Acids Res 43(6):e39
	19	Krumm, 2012, Copy number variation detection and genotyping from exome sequence data, Genome Res 22 (8):1525-1532
	20	Lai, 2004, Characterization of the maize endosperm transcriptome and its comparison to the rice genome, Genome Res14:1932-1937
	21	Langmead, 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol 10:R25
	22	Li, 2012, CONTRA: copy number analysis for targeted resequencing, Bioinformatics 28(10):1307-1313
	23	Lindstrom, 2004, Pyrosequencing for detection of Lamivudine-resistant Hepatitis B virus, J Clin Microb 42 (10):4788-4795
	24	Liu, 2008, Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples, BMC Genomics 9(509):1-17
	25	Ma, 2015, Quantitative Analysis of Copy Number Variants Based on Real-Time LightCycler PCR, Curr Protoc Hum Genet 80:7.21.1-7.23.8
	26	Machine translation generated on March 7, 2018, of CN 105890722 by website of European Patent Office (4 pages)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	27	Margulies, 2005, Genorne sequencing in open microfabricated high density picoliter reactors, Nature 437 (7057):376-380
	28	McCloskey, 2007, Encoding PCR products with batch-stamps and barcodes, Biochem Genet 45:761-767
	29	Miner, 2004, Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR, Nucl Acids Res 32 (17):e135
	30	Myers, 2013, Protocol for Creating Multiplexed miRNA Libraries for Use in Illumina Sequencing, Myers lab microRNA-seq Protocol, Hudson Alpha Institute for Biotechnology web site, dated May 2, 2013, (15 pages)
	31	NuGEN, 2014, User Guide Ovation Target Enrichment System, NuGEN Technologies Inc., San Carlos, CA (45 pages)
	32	Plagnol, 2012, A robust model for read count data in exome sequencing experiments and implications for copy number variant calling, Bioinformatics 28(21):2747-2754
	33	Qiu, 2003, DNA sequence-based "bar-codes" for tracking the origins of expressed sequence tags from a maize cDNA library constructed using multiple mRNA sources, Plant Physiol 133:475-481
	34	Querfurth, 2012, Creation and application of immortalized bait libraries for targeted enrichment and next-generation sequencing, Biotechniques 52(6):375-380
	35	Ronaghi, 2001, Pyrosequencing sheds light on DNA sequencing, Genome Res 11:3-11
	36	Sathirapongsasuti, 2011, Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV, Bioinformatics 27(19):2648-2654
	37	Schiemer, 2011, Illumina TruSeq Adapters Demystified,TUFTS UNIVERSITY CORE FACILITY XP055357867 (5 pages)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	38	Shapero, 2001, SNP Genotyping by multiplexed solid-phase amplification and fluorescent minisequencing, Genome Res 11:1926-1934
	39	Shendure, 2005, Accurate multiplex polony sequencing of an evolved bacterial genome, Science 309:1728
	40	Soni, 2007, Progress toward ultrafast DNA sequencing using solid-state nanopores, Clin Chem 53(11):1996-2001
	41	Sood, 2006, Methods for reverse genetic screening in zebrafish by resequencing and TILLING, Methods 39:220-227
	42	Staroscik, 2004, Calculator for determining the number of copies of a template, URI Genomics, webpage archive dated April 6, 2017 (1 page), Retreived from the internet on March 7, 2018, from <https://web.archive.org/web/20170406174850/http://cels.uri.edu/gsc/cndna.html>
	43	Stratagene, 1998, Gene characterization kits, Stratagene Catalog, p. 39 (2 pages)
	44	Supplementary European search report and opinion mailed January 30, 2018, for European patent application No. 15830393.3 (6 pages)
	45	Till, 2003, Large-scale discovery of induced point mutations with high-throughput TILLING, Genome Res 13:524-530
	46	Trapnell, 2010, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat Biotech 28:511-515
	47	Trapnell, 2013, Differential analysis of gene regulation at transcript resolution with RNA-seq, Nat Biotech 31:46-53
	48	Unemo, 2004, Molecular typing of Neisseria gonorrhoeae isolates by pyrosequencing of highly polymorphic segments of the porB gene, J Clin Microb 42(7):2926-2934

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0395

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	49	Vigal, 2002, A review on SNP and other types of molecular markers and their use in animal genetics, Genet Sel Evol 34:275-305
	50	Walker, 1992, Strand displacement amplification–an isothermal, in vitro DNA amplification technique, Nucl Acids Res 20(7):1691-1696

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/TERESA E STRZELECKA/	Date Considered	12/02/2020

*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. /T.E.S/

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.
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516	3249

21710 7590 12/08/2020

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
STRZELECKA, TERESA E	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
12/08/2020	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):

ip@brownrudnick.com
usactions@brownrudnick.com

IDT Exhibit 1019.0397

Office Action Summary	Application No. 16/743,604	Applicant(s) Amorese et al.	
	Examiner TERESA E STRZELECKA	Art Unit 1637	AIA (FITF) Status Yes

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

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

☐ 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) 22-32 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) 22-32 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 01/15/20 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 7/22/20.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

IDT Exhibit 1019.0398

DETAILED ACTION

Notice of Pre-AIA or AIA Status

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

2. Claims 22-32 are pending and under consideration.

Information Disclosure Statement

3. The information disclosure statement (IDS) submitted on July 22, 2020 is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Claim Interpretation

4. Applicant defined the term "adaptor" in paragraph [0027] as follows:

"An adapter refers to an oligonucleotide sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest."

Therefore, since any oligonucleotide can perform such function, any oligonucleotide is considered to anticipate this term.

5. Applicant defined the term "identifier site" in paragraph [0030] as follows:

"An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads."

6. Applicant defined the term "universal primer" in paragraph [0035] as follows:

"A universal primer is a primer that can be used for amplifying a number of different sequences."

7. Applicant did not define the term "junction", therefore it is considered as any region of the adapter.

Claim Rejections - 35 USC § 103

8. 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.

9. The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

10. This application currently names joint inventors. In considering patentability of the claims the examiner presumes that the subject matter of the various claims was commonly owned as of the effective filing date of the claimed invention(s) absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and effective filing dates of each claim that was not commonly owned as of the effective filing date of the later invention in order for the examiner to consider the applicability of 35 U.S.C. 102(b)(2)(C) for any potential 35 U.S.C. 102(a)(2) prior art against the later invention.

11. Claims 22-31 are rejected under 35 U.S.C. 103 as being unpatentable over Samuels et al. (US 2012/0220494 A1; filed February 16, 2012 and published August 30, 2012; cited in the IDS)

and McCloskey et al. (Biochem. Genet., vol. 45, pp. 761-767, 2007; cited in the IDS), as evidenced by Illumina publication No. 770-2008-011 (pp. 1-4, published December 2, 2008).

A) Regarding claims 22, 29 and 30, Samuels et al. teach a method for detecting duplicate sequencing reads, the method comprising:

ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides and a primer binding site ([0205]; [0209], [0212], [0216]-[0217]; Fig. 6; [0231]-[0235]; [0245]; where the adaptors have an identifier site between 4 and 20 bp and a primer binding site);

amplifying the adaptor-ligated fragments into amplicons ([0288]-[0289]; [0326]-[0327]);

sequencing the amplicons to produce sequence reads from amplified that include identifier and target sequences ([0326]-[0328]; [0328]-[0337]); and

identifying sequence reads with identical identifier site and target sequence ([0328]; [0339]-[0341]; Fig. 43 and 44; where sequencing errors due to PCR are detected).

Regarding claim 24, Samuels et al. teach sequencing the identifier site with the indexing site (Fig. 43; [0339]-[0341]).

Regarding claim 25, Samuels et al. teach Illumina sequencing ([0028]; [0295]; [0403]). As evidenced by Illumina sequencing brochure, the process involves separately sequencing the identifier and the target (Fig. 1; page 2, paragraphs 5-7). Therefore by teaching Illumina sequencing Samuels et al. inherently teach separately sequencing the target and identifier.

Regarding claims 26 and 27, Samuels et al. teach DNA and RNA fragments, and genomic DNA (page 2, [0017]).

Regarding claim 28, Samuels et al. teach target sequence binding site being a universal binding site ([0216]).

Regarding claim 31, Samuels et al. teach the identifier being at a junction between the adapter and nucleic acid fragment (Fig. 6 and 7, for example)

B) Samuels et al. teach detection of PCR-introduced errors in sequencing, but do not specifically teach detecting duplicate sequencing reads.

C) Regarding claim 22 and 23, McCloskey et al. teach attaching an adaptor containing an identifier site and a target sequence binding site to genomic DNA by amplification, PCR-amplifying the resulting construct and sequencing (Fig. 1; page 762, paragraphs 3-5; page 763; page 764, paragraphs 1-4).

Regarding claims 22 and 23, McCloskey et al. teach sequencing the PCR products containing the adapter sequences, grouping the sequencing reads based on the batch stamp and barcode and detecting and removing duplicate reads (page 766, fourth and fifth paragraphs).

Regarding claims 29 and 30, McCloskey et al. teach indexing site between six and eight bp (page 764, second paragraph).

It would have been *prima facie* obvious to one of ordinary skill in the art at the time of the invention to have used the batch-stamp and barcode-encoded oligonucleotides as suggested by McCloskey et al. in the method of Samuels et al. to eliminate PCR-originating artefacts. As stated by McCloskey et al. (Abstract):

"Polymerase chain reaction (PCR) has become the mainstay of DNA sequence analysis. Yet there is always uncertainty concerning the source of the template DNA that gave rise to a particular PCR product. The risks of contamination, biased amplification, and product redundancy are especially high when limited amounts of template DNA are used. We have developed and applied molecular encoding principles to solve this source-uncertainty problem for DNA sequences generated by standard PCR. Batch-stamps specify the date and sample

identity, and barcodes detect template redundancy. Our approach thus enables classification of each PCR-derived sequence as valid, contaminant, or redundant, and provides a measure of sequence diversity. We recommend that batch-stamps and barcodes be used when amplifying irreplaceable DNAs and cDNAs available for forensic, clinical, single cell, and ancient DNA analyses."

12. Claim 32 is rejected under 35 U.S.C. 103 as being unpatentable over Samuels et al. (US 2012/0220494 A1; filed February 16, 2012 and published August 30, 2012; cited in the IDS) and McCloskey et al. (Biochem. Genet., vol. 45, pp. 761-767, 2007; cited in the IDS), as evidenced by Illumina publication No. 770-2008-011 (pp. 1-4, published December 2, 2008), as applied to claim 22 above, and further in view of Ule et al. (US 2011/0269647 A1; published November 2011) and Vasli et al. (Acta Neuropathol., vol. 124, pp. 273-283 plus supplemental material pp. 1-10, 2012).

A) Regarding claim 32, neither reference specifically teaches determination of the genomic start position of each read.

B) Regarding claim 32, Ule et al. teach a method of identifying sequences on RNA which interact with RNA-binding proteins by next generation sequencing using barcoded adapters ([0007]; [0017]-[0028]; [0039]; Fig. 2; [0041]; [0082]-[0113]; [0192]-[0196]). Ule et al. teach separation of sequencing reads based on barcodes and removing duplicates which are molecules which contain the same start cDNA sequence ([0196]; [0200]).

Regarding claim 32, Vasli et al. teach determination of mutations associated with neurodegenerative diseases by next generation sequencing using Illumina sequencing protocol (page 274, paragraphs 5-6; page 275, first paragraph). Vasli et al. teach removing duplicate reads sharing the same start position and strand (page 275, second paragraph).

It would have been *prima facie* obvious to one of ordinary skill in the art before the effective filing date of the claimed invention to have used the start position and barcode as suggested by Ule et al. and Vasli et al. to remove duplicate sequence reads in the method of Samuels et al. and McClosky et al. The motivation to do so would have been that reads containing both the same barcode and same genomic start position would be more probable to be true duplicate reads.

Double Patenting

13. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on nonstatutory double patenting provided the reference application or patent either is shown to be commonly owned with the examined application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. See MPEP § 717.02 for applications subject to examination

under the first inventor to file provisions of the AIA as explained in MPEP § 2159. See MPEP § 2146 *et seq.* for applications not subject to examination under the first inventor to file provisions of the AIA. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

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

14. Claims 22, 23 and 26-31 are rejected on the ground of nonstatutory double patenting as being unpatentable over claim 1, 2, 9, 10, 13, 14 and 16 of U.S. Patent No. 9,546,399. Although the claims at issue are not identical, they are not patentably distinct from each other because claims of the '399 patent anticipate instant claims.

Specifically, claim 22 of the instant application is drawn to a method for detecting duplicate sequencing reads, the method comprising:

ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides and a primer binding site;

amplifying the adaptor-ligated fragments into amplicons;

sequencing the amplicons to produce sequence reads that include identifier and target sequences; and

identifying sequence reads with identical identifier and target sequences as duplicates.

Claim 1 of the '399 patent is drawn to a method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:

a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:

(i) an indexing primer binding site;

(ii) an indexing site;

(iii) an identifier site consisting of between 1 and 8 nucleotides; and

(iv) a target sequence primer binding site;

b) amplifying the adapter-nucleic acid fragment ligated products;

c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and

d) identifying a duplicate sequencing read as the sequencing read comprising the same identifier site and nucleic acid fragment as another sequencing read in the population of sequencing reads;

wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is variable in sequence content in a plurality of adaptors.

Therefore claim 1 of the '399 patent anticipates instant claims 22 and 31. The dependent claims 23 and 26-30 are anticipated by claims 2, 9, 10, 16, 13 and 14 of the '399 patent.

15. Claim 32 is rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1, 2, 9, 10, 13, 14 and 16 of U.S. Patent No. 9,546,399 in view of Ule et al. (US 2011/0269647 A1; published November 2011) and Vasli et al. (Acta Neuropathol., vol. 124, pp. 273-283 plus supplemental material pp. 1-10, 2012).

A) Regarding claim 32, claims of the '399 patent do not teach determination of the genomic start position of each read.

B) Regarding claim 32, Ule et al. teach a method of identifying sequences on RNA which interact with RNA-binding proteins by next generation sequencing using barcoded adapters ([0007]; [0017]-[0028]; [0039]; Fig. 2; [0041]; [0082]-[0113]; [0192]-[0196]). Ule et al. teach seaparation of sequencing reads based on barcodes and removing duplicates which are molecules which contain the same start cDNA sequence ([0196]; [0200]).

Regarding claim 32, Vasli et al. teach determination of mutations associated with neurodegenerative diseases by next generation sequencing using Illumina sequencing protocol (page 274, paragraphs 5-6; page 275, first paragraph). Vasli et al. teach removing duplicate reads sharing the same start position and strand (page 275, second paragraph).

It would have been *prima facie* obvious to one of ordinary skill in the art before the effective filing date of the claimed invention to have used the start position and barcode as suggested by Ule et al. and Vasli et al. to remove duplicate sequence reads in the method of the claims of the '399 patent. The motivation to do so would have been that reads containing both the same barcode and same genomic start position would be more probable to be true duplicate reads.

16. No claims are allowed.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to TERESA E STRZELECKA whose telephone number is (571)272-0789. The examiner can normally be reached on M-F 9-7.

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

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, 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 https://ppair-my.uspto.gov/pair/PrivatePair. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/TERESA E STRZELECKA/
Primary Examiner, Art Unit 1637
December 3, 2020

Web Search History

date, time	web site	search string
12/2/2020 5:28:42 PM	Google	MPEP 713.01
12/2/2020 5:58:19 PM	Google	genomic start position for a sequencing read
12/2/2020 5:59:19 PM	Google	"genomic start position" "sequencing read"
12/2/2020 6:05:59 PM	Google	"genomic start position" sequencing
12/2/2020 6:15:12 PM	Google	"genomic start position" sequencing alignment
12/2/2020 6:25:59 PM	Google	"start position" sequencing alignment
12/2/2020 6:28:12 PM	Google	"start position" next generation sequencing alignment
12/2/2020 7:44:54 PM	Google	"start position" next generation sequencing barcode

IDT Exhibit 1019.0409

WEST Search History for Application 16743604

Creation Date: 2020120216:15

Prior Art Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
("20010000077" | "20010031739" | "20010034048" | "20010041334" | "20020028447" | "20020058270" | "20020115088" | "20020150919" | "20020155451" | "20020164628" | "20020164634" | "20020197639" | "20030017591" | "20030022207" | "20030082543" | "20030087251" | "20030119150" | "20030143555" | "20030175780" | "20030180779" | "20030186234" | "20030207279" | "20030211616" | "20030215926" | "20030224439" | "20030232348" | "20040002371" | "20040005614" | "20040023271" | "20040115815" | "20040137456" | "20040161742" | "20040203019" | "20040203025" | "20040248153" | "20050003441" | "20050014192" | "20050019793" | "20050059048" | "20050064456" | "20050123956" | "20050136417" | "20050142577" | "20050191656" | "20050191682" | "20050208538" | "20060008824" | "20060014182" | "20060024678" | "20060024711" | "20060035274" | "20060046251" | "20060051789" | "20060068415" | "20060134633" | "20060216724" | "20060263789" | "20060281082" | "20060286566" | "20060292597" | "20070031857" | "20070134128" | "20070141604" | "20070224607" | "20070224613" | "20070231823" | "20070238122" | "20070263045" | "20080038727" | "20080087826" | "20080103058" | "20080131937" | "20080160580" | "20080176311" | "20080182300" | "20080194413" | "20080194416" | "20080206764" | "20080213770" | "20080217246" | "20080241831" | "20080242560" |	PGPB, USPT	n/a	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0410

"20080286795" | "20090011959" |						
"20090024331" | "20090026082" |						
"20090029385" | "20090036663" |						
"20090061425" | "20090061439" |						
"20090068645" | "20090068655" |						
"20090068709" | "20090105081" |						
"20090117573" | "20090117621" |						
"20090124514" | "20090127589" |						
"20090130721" | "20090203085" |						
"20090203531" | "20090233802" |						
"20090233804" | "20090239232" |						
"20090275486" | "20090280538" |						
"20090298075" | "20100015666" |						
"20100021973" | "20100022403" |						
"20100029511" | "20100105052" |						
"20100113296" | "20100129879" |						
"20100137143" | "20100159559" |						
"20100167954" | "20100173394" |						
"20100203597" | "20100267043" |						
"20100273219" | "20100311066" |						
"20100323348" | "20110015096" |						
"20110039732" | "20110104785" |						
"20110105364" | "20110129827" |						
"20110189679" | "20110224105" |						
"20110288780" | "20110294132" |						
"20110319290" | "20120003657" |						
"20120028310" | "20120045797" |						
"20120071331" | "20120074925" |						
"20120102054" | "20120107811" |						
"20120122701" | "20120149068" |						
"20120156728" | "20120157322" |						
"20120190587" | "20120208705" |						
"20120220483" | "20120220494" |						
"20120237943" | "20120238738" |						
"20120245041" | "20120270212" |						
"20120283145" | "20120289426" |						
"20120309002" | "20130005585" |						
"20130059738" | "20130231253" |						
"20140038188" | "20140038236" |						
"20140065692" | "20140274729" |						
"20140274731" | "20140274738" |						
"20140303000" | "20140378345" |						
"20150011396" | "20150017635" |						
"20150101595" | "20150133319" |						
"20150284769" | "20150299767" |						
"20150299784" | "20150299812" |						
"20160122756" | "20160130576" |						
"20160153039" | "20160203259" |						
"20160220994" | "20160251711" |						
"20160265042" | "20160275240" |						
"20160296930" | "4362867" | "4458066" |						

IDT Exhibit 1019.0411

"4469863"	"4582877"	"4876187"						
"4925065"	"4935357"	"4942124"						
"4988617"	"4996143"	"5011769"						
"5034506"	"5035996"	"5043272"						
"5082830"	"5090591"	"5130238"						
"5169766"	"5171534"	"5194370"						
"5216141"	"5234809"	"5235033"						
"5242794"	"5384242"	"5386023"						
"5399491"	"5409818"	"5418149"						
"5422271"	"5427929"	"5480784"						
"5494810"	"5508169"	"5508178"						
"5510270"	"5525471"	"5545522"						
"5554516"	"5554517"	"5556752"						
"5565340"	"5573913"	"5578832"						
"5589339"	"5602240"	"5637684"						
"5641658"	"5644048"	"5665549"						
"5667976"	"5667979"	"5679512"						
"5681726"	"5683879"	"5688648"						
"5705628"	"5708154"	"5710028"						
"5712126"	"5716785"	"5726329"						
"5750341"	"5759822"	"5763178"						
"5789206"	"5824517"	"5824518"						
"5837832"	"5876976"	"5882867"						
"5888779"	"5888819"	"5945313"						
"5952176"	"5958681"	"5965409"						
"5969119"	"5972618"	"6004744"						
"6004745"	"6027889"	"6027923"						
"6030774"	"6037152"	"6056661"						
"6077674"	"6087103"	"6090553"						
"6090591"	"6107023"	"6110709"						
"6150112"	"6159685"	"6160105"						
"6169194"	"6172208"	"6174680"						
"6180338"	"6190865"	"6194211"						
"6197501"	"6197557"	"6210891"						
"6225109"	"6225451"	"6232104"						
"6251639"	"6262490"	"6270961"						
"6280935"	"6287766"	"6287825"						
"6291170"	"6306365"	"6306597"						
"6309843"	"6326142"	"6335167"						
"6339147"	"6440705"	"6449562"						
"6582938"	"6670461"	"6686156"						
"6692918"	"6770748"	"6777180"						
"6794499"	"6815164"	"6815167"						
"6825011"	"6833246"	"6849404"						
"6858413"	"6913884"	"6917726"						
"6924104"	"6946251"	"7001724"						
"7033764"	"7048481"	"7052847"						
"7056676"	"7056716"	"7060441"						
"7094536"	"7115400"	"7170050"						
"7175982"	"7176025"	"7189512"						
"7211390"	"7232656"	"7244559"						

"7244567" | "7264929" | "7273730" | "7276720" | "7294461" | "7300755" | "7302146" | "7313308" | "7315019" | "7323305" | "7335762" | "7351557" | "7354717" | "7361466" | "7361468" | "7402386" | "7405281" | "7414117" | "7416844" | "7462452" | "7462468" | "7476503" | "7476504" | "7491498" | "7501245" | "7579153" | "7704687" | "7741463" | "7771934" | "7771946" | "7803550" | "7846666" | "7846733" | "7867703" | "7939258" | "7948015" | "7985565" | "8017335" | "8034568" | "8053192" | "8071311" | "8143001" | "8209130" | "8334116" | "8465950" | "8492095" | "8512956" | "8551709" | "8852867" | "8999677" | "9175325" | "9175336" | "9181582" | "9206418" | "9248076" | "9546399" | "9702004" | "9745627" | "9896721").PN.						
("4362867" | "4458066" | "4469863" | "4582877" | "4876187" | "4925065" | "4935357" | "4942124" | "4988617" | "4996143" | "5011769" | "5034506" | "5035996" | "5043272" | "5082830" | "5090591" | "5130238" | "5169766" | "5171534" | "5194370" | "5216141" | "5234809" | "5235033" | "5242794" | "5384242" | "5386023" | "5399491" | "5409818" | "5418149" | "5422271" | "5427929" | "5480784" | "5494810" | "5508169" | "5508178" | "5510270" | "5525471" | "5545522" | "5554516" | "5554517" | "5556752" | "5565340" | "5573913" | "5578832" | "5589339" | "5602240" | "5637684" | "5641658" | "5644048" | "5665549" | "5667976" | "5667979" | "5679512" | "5681726" | "5683879" | "5688648" | "5705628" | "5708154" | "5710028" | "5712126" | "5716785" | "5726329" | "5750341" | "5759822" | "5763178" | "5789206" | "5824517" | "5824518" | "5837832" | "5876976" | "5882867" | "5888779" | "5888819" | "5945313" | "5952176" | "5958681" | "5965409" | "5969119" | "5972618" | "6004744" | "6004745" | "6027889" | "6027923" | "6030774" | "6037152" | "6056661" | "6077674" | "6087103" | "6090553" | "6090591" |	USPT	214	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0413

"6107023" | "6110709" | "6150112" | "6159685" | "6160105" | "6169194" | "6172208" | "6174680" | "6180338" | "6190865" | "6194211" | "6197501" | "6197557" | "6210891" | "6225109" | "6225451" | "6232104" | "6251639" | "6262490" | "6270961" | "6280935" | "6287766" | "6287825" | "6291170" | "6306365" | "6306597" | "6309843" | "6326142" | "6335167" | "6339147" | "6440705" | "6449562" | "6582938" | "6670461" | "6686156" | "6692918" | "6770748" | "6777180" | "6794499" | "6815164" | "6815167" | "6825011" | "6833246" | "6849404" | "6858413" | "6913884" | "6917726" | "6924104" | "6946251" | "7001724" | "7033764" | "7048481" | "7052847" | "7056676" | "7056716" | "7060441" | "7094536" | "7115400" | "7170050" | "7175982" | "7176025" | "7189512" | "7211390" | "7232656" | "7244559" | "7244567" | "7264929" | "7273730" | "7276720" | "7294461" | "7300755" | "7302146" | "7313308" | "7315019" | "7323305" | "7335762" | "7351557" | "7354717" | "7361466" | "7361468" | "7402386" | "7405281" | "7414117" | "7416844" | "7462452" | "7462468" | "7476503" | "7476504" | "7491498" | "7501245" | "7579153" | "7704687" | "7741463" | "7771934" | "7771946" | "7803550" | "7846666" | "7846733" | "7867703" | "7939258" | "7948015" | "7985565" | "8017335" | "8034568" | "8053192" | "8071311" | "8143001" | "8209130" | "8334116" | "8465950" | "8492095" | "8512956" | "8551709" | "8852867" | "8999677" | "9175325" | "9175336" | "9181582" | "9206418" | "9248076" | "9546399" | "9702004" | "9745627" | "9896721").PN.						
(("4362867" | "4458066" | "4469863" | "4582877" | "4876187" | "4925065" | "4935357" | "4942124" | "4988617" | "4996143" | "5011769" | "5034506" | "5035996" | "5043272" | "5082830" | "5090591" | "5130238" | "5169766" | "5171534" | "5194370" | "5216141" | "5234809" | "5235033" | "5242794" |	USPT	2	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0414

"5384242" | "5386023" | "5399491" |						
"5409818" | "5418149" | "5422271" |						
"5427929" | "5480784" | "5494810" |						
"5508169" | "5508178" | "5510270" |						
"5525471" | "5545522" | "5554516" |						
"5554517" | "5556752" | "5565340" |						
"5573913" | "5578832" | "5589339" |						
"5602240" | "5637684" | "5641658" |						
"5644048" | "5665549" | "5667976" |						
"5667979" | "5679512" | "5681726" |						
"5683879" | "5688648" | "5705628" |						
"5708154" | "5710028" | "5712126" |						
"5716785" | "5726329" | "5750341" |						
"5759822" | "5763178" | "5789206" |						
"5824517" | "5824518" | "5837832" |						
"5876976" | "5882867" | "5888779" |						
"5888819" | "5945313" | "5952176" |						
"5958681" | "5965409" | "5969119" |						
"5972618" | "6004744" | "6004745" |						
"6027889" | "6027923" | "6030774" |						
"6037152" | "6056661" | "6077674" |						
"6087103" | "6090553" | "6090591" |						
"6107023" | "6110709" | "6150112" |						
"6159685" | "6160105" | "6169194" |						
"6172208" | "6174680" | "6180338" |						
"6190865" | "6194211" | "6197501" |						
"6197557" | "6210891" | "6225109" |						
"6225451" | "6232104" | "6251639" |						
"6262490" | "6270961" | "6280935" |						
"6287766" | "6287825" | "6291170" |						
"6306365" | "6306597" | "6309843" |						
"6326142" | "6335167" | "6339147" |						
"6440705" | "6449562" | "6582938" |						
"6670461" | "6686156" | "6692918" |						
"6770748" | "6777180" | "6794499" |						
"6815164" | "6815167" | "6825011" |						
"6833246" | "6849404" | "6858413" |						
"6913884" | "6917726" | "6924104" |						
"6946251" | "7001724" | "7033764" |						
"7048481" | "7052847" | "7056676" |						
"7056716" | "7060441" | "7094536" |						
"7115400" | "7170050" | "7175982" |						
"7176025" | "7189512" | "7211390" |						
"7232656" | "7244559" | "7244567" |						
"7264929" | "7273730" | "7276720" |						
"7294461" | "7300755" | "7302146" |						
"7313308" | "7315019" | "7323305" |						
"7335762" | "7351557" | "7354717" |						
"7361466" | "7361468" | "7402386" |						
"7405281" | "7414117" | "7416844" |						
"7462452" | "7462468" | "7476503" |						

"7476504" | "7491498" | "7501245" | "7579153" | "7704687" | "7741463" | "7771934" | "7771946" | "7803550" | "7846666" | "7846733" | "7867703" | "7939258" | "7948015" | "7985565" | "8017335" | "8034568" | "8053192" | "8071311" | "8143001" | "8209130" | "8334116" | "8465950" | "8492095" | "8512956" | "8551709" | "8852867" | "8999677" | "9175325" | "9175336" | "9181582" | "9206418" | "9248076" | "9546399" | "9702004" | "9745627" | "9896721").PN.) and sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8)						
("20010000077" | "20010031739" | "20010034048" | "20010041334" | "20020028447" | "20020058270" | "20020115088" | "20020150919" | "20020155451" | "20020164628" | "20020164634" | "20020197639" | "20030017591" | "20030022207" | "20030082543" | "20030087251" | "20030119150" | "20030143555" | "20030175780" | "20030180779" | "20030186234" | "20030207279" | "20030211616" | "20030215926" | "20030224439" | "20030232348" | "20040002371" | "20040005614" | "20040023271" | "20040115815" | "20040137456" | "20040161742" | "20040203019" | "20040203025" | "20040248153" | "20050003441" | "20050014192" | "20050019793" | "20050059048" | "20050064456" | "20050123956" | "20050136417" | "20050142577" | "20050191656" | "20050191682" | "20050208538" | "20060008824" | "20060014182" | "20060024678" | "20060024711" | "20060035274" | "20060046251" | "20060051789" | "20060068415" | "20060134633" | "20060216724" | "20060263789" | "20060281082" | "20060286566" | "20060292597" | "20070031857" | "20070134128" | "20070141604" | "20070224607" | "20070224613" | "20070231823" | "20070238122" | "20070263045" |	PGPB	181	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0416

"20080038727"	"20080087826"						
"20080103058"	"20080131937"						
"20080160580"	"20080176311"						
"20080182300"	"20080194413"						
"20080194416"	"20080206764"						
"20080213770"	"20080217246"						
"20080241831"	"20080242560"						
"20080286795"	"20090011959"						
"20090024331"	"20090026082"						
"20090029385"	"20090036663"						
"20090061425"	"20090061439"						
"20090068645"	"20090068655"						
"20090068709"	"20090105081"						
"20090117573"	"20090117621"						
"20090124514"	"20090127589"						
"20090130721"	"20090203085"						
"20090203531"	"20090233802"						
"20090233804"	"20090239232"						
"20090275486"	"20090280538"						
"20090298075"	"20100015666"						
"20100021973"	"20100022403"						
"20100029511"	"20100105052"						
"20100113296"	"20100129879"						
"20100137143"	"20100159559"						
"20100167954"	"20100173394"						
"20100203597"	"20100267043"						
"20100273219"	"20100311066"						
"20100323348"	"20110015096"						
"20110039732"	"20110104785"						
"20110105364"	"20110129827"						
"20110189679"	"20110224105"						
"20110288780"	"20110294132"						
"20110319290"	"20120003657"						
"20120028310"	"20120045797"						
"20120071331"	"20120074925"						
"20120102054"	"20120107811"						
"20120122701"	"20120149068"						
"20120156728"	"20120157322"						
"20120190587"	"20120208705"						
"20120220483"	"20120220494"						
"20120237943"	"20120238738"						
"20120245041"	"20120270212"						
"20120283145"	"20120289426"						
"20120309002"	"20130005585"						
"20130059738"	"20130231253"						
"20140038188"	"20140038236"						
"20140065692"	"20140274729"						
"20140274731"	"20140274738"						
"20140303000"	"20140378345"						
"20150011396"	"20150017635"						
"20150101595"	"20150133319"						

IDT Exhibit 1019.0417

"20150284769" | "20150299767" | "20150299784" | "20150299812" | "20160122756" | "20160130576" | "20160153039" | "20160203259" | "20160220994" | "20160251711" | "20160265042").PN.						
(("20010000077" | "20010031739" | "20010034048" | "20010041334" | "20020028447" | "20020058270" | "20020115088" | "20020150919" | "20020155451" | "20020164628" | "20020164634" | "20020197639" | "20030017591" | "20030022207" | "20030082543" | "20030087251" | "20030119150" | "20030143555" | "20030175780" | "20030180779" | "20030186234" | "20030207279" | "20030211616" | "20030215926" | "20030224439" | "20030232348" | "20040002371" | "20040005614" | "20040023271" | "20040115815" | "20040137456" | "20040161742" | "20040203019" | "20040203025" | "20040248153" | "20050003441" | "20050014192" | "20050019793" | "20050059048" | "20050064456" | "20050123956" | "20050136417" | "20050142577" | "20050191656" | "20050191682" | "20050208538" | "20060008824" | "20060014182" | "20060024678" | "20060024711" | "20060035274" | "20060046251" | "20060051789" | "20060068415" | "20060134633" | "20060216724" | "20060263789" | "20060281082" | "20060286566" | "20060292597" | "20070031857" | "20070134128" | "20070141604" | "20070224607" | "20070224613" | "20070231823" | "20070238122" | "20070263045" | "20080038727" | "20080087826" | "20080103058" | "20080131937" | "20080160580" | "20080176311" | "20080182300" | "20080194413" | "20080194416" | "20080206764" | "20080213770" | "20080217246" | "20080241831" | "20080242560" | "20080286795" | "20090011959" | "20090024331" | "20090026082" | "20090029385" | "20090036663" |	PGPB	5	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0418

"20090061425" | "20090061439" | "20090068645" | "20090068655" | "20090068709" | "20090105081" | "20090117573" | "20090117621" | "20090124514" | "20090127589" | "20090130721" | "20090203085" | "20090203531" | "20090233802" | "20090233804" | "20090239232" | "20090275486" | "20090280538" | "20090298075" | "20100015666" | "20100021973" | "20100022403" | "20100029511" | "20100105052" | "20100113296" | "20100129879" | "20100137143" | "20100159559" | "20100167954" | "20100173394" | "20100203597" | "20100267043" | "20100273219" | "20100311066" | "20100323348" | "20110015096" | "20110039732" | "20110104785" | "20110105364" | "20110129827" | "20110189679" | "20110224105" | "20110288780" | "20110294132" | "20110319290" | "20120003657" | "20120028310" | "20120045797" | "20120071331" | "20120074925" | "20120102054" | "20120107811" | "20120122701" | "20120149068" | "20120156728" | "20120157322" | "20120190587" | "20120208705" | "20120220483" | "20120220494" | "20120237943" | "20120238738" | "20120245041" | "20120270212" | "20120283145" | "20120289426" | "20120309002" | "20130005585" | "20130059738" | "20130231253" | "20140038188" | "20140038236" | "20140065692" | "20140274729" | "20140274731" | "20140274738" | "20140303000" | "20140378345" | "20150011396" | "20150017635" | "20150101595" | "20150133319" | "20150284769" | "20150299767" | "20150299784" | "20150299812" | "20160122756" | "20160130576" | "20160153039" | "20160203259" | "20160220994" | "20160251711" | "20160265042").PN.) and sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8)						

IDT Exhibit 1019.0419

16/743604.app.	PGPB	1	ADJ	YES	ASSIGNEE	12-02-2020
sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8) and (duplicat$5 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	PGPB	47	ADJ	YES	ASSIGNEE	12-02-2020
sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8) and (duplicat$5 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	USPT	17	ADJ	YES	ASSIGNEE	12-02-2020
sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8) and (duplicat$5 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	DWPI	2	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0420

Notice of References Cited	Application/Control No. 16/743,604	Applicant(s)/Patent Under Reexamination Amorese et al.	
	Examiner TERESA E STRZELECKA	Art Unit 1637	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-20110269647-A1	11-2011	ULE; JERNEJ	C12N15/1096	506/26
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Illumina publication No. 770-2008-011, Multiplexed Sequencing with the Illumina Genome Analyzer System, pp. 1-4 (Year: 2008)
	V	Vasli et al., Next generation sequencing for molecular diagnosis of neuromuscular diseases, Acta Neuropathol., vol. 124, pp. 273-283 plus supplemental material pp. 1-10 (Year: 2012)
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

IDT Exhibit 1019.0421

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	16743604
Filing Date	2020-01-15
First Named Inventor	Douglas A. Amorese
Art Unit	1637
Examiner Name	T. E. Strzelecka
Attorney Docket Number	NUGN-001/03US 34079/516

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	20090029385	A1	2009-01-29	Christians et al.	
	2	20090036663	A1	2009-02-05	Kurn	
	3	20090061425	A1	2009-03-05	Lo et al.	
	4	20090061439	A1	2009-03-05	Buzby	
	5	20090068645	A1	2009-03-12	Sibson	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	6	20090068655	A1	2009-03-12	Williams	
	7	20090068709	A1	2009-03-12	Kurn et al.	
	8	20090105081	A1	2009-04-23	Rodesch et al.	
	9	20090117573	A1	2009-05-07	Fu et al.	
	10	20090117621	A1	2009-05-07	Boutell et al.	
	11	20090124514	A1	2009-05-14	Fu et al.	
	12	20090127589	A1	2009-05-21	Rothberg et al.	
	13	20090130721	A1	2009-05-21	Kurn et al.	
	14	20090203085	A1	2009-08-13	KURN et al.	
	15	20090203531	A1	2009-08-13	Kurn	
	16	20090233802	A1	2009-09-17	Bignell et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	17	20090233804	A1	2009-09-17	Kurn et al.	
	18	20090239232	A1	2009-09-24	Kurn	
	19	20090275486	A1	2009-11-05	Kurn et al.	
	20	20090280538	A1	2009-11-12	Patel et al.	
	21	20090298075	A1	2009-12-03	Travers et al.	
	22	20100015666	A1	2010-01-21	Brenner et al.	
	23	20100021973	A1	2010-01-28	Makarov et al.	
	24	20100022403	A1	2010-01-28	Kurn et al.	
	25	20100029511	A1	2010-02-04	Raymond et al.	
	26	20100105052	A1	2010-04-29	DRMANAC et al.	
	27	20100113296	A1	2010-05-06	Myerson	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0424

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	28	20100129879	A1	2010-05-27	Ach et al.	
	29	20100137143	A1	2010-06-03	Rothberg et al.	
	30	20100159559	A1	2010-06-24	Kurn et al.	
	31	20100167954	A1	2010-07-01	Earnshaw et al.	
	32	20100173394	A1	2010-07-08	Colston, JR. et al.	
	33	20100203597	A1	2010-08-12	Chen et al.	
	34	20100267043	A1	2010-10-21	Braverman et al.	
	35	20100273219	A1	2010-10-28	May et al.	
	36	20100311066	A1	2010-12-09	Kurn	
	37	20100323348	A1	2010-12-23	Hamady et al.	
	38	20110015096	A1	2011-01-20	CHIU	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	39	20110039732	A1	2011-02-17	Raymond et al.	
	40	20110104785	A1	2011-05-05	Vaidyanathan et al.	
	41	20110105364	A1	2011-05-05	Kurn	
	42	20110129827	A1	2011-06-02	Causey et al.	
	43	20110189679	A1	2011-08-04	Kurn et al.	
	44	20110224105	A1	2011-09-15	Kurn et al.	
	45	20110288780	A1	2011-11-24	Rabinowitz et al.	
	46	20110294132	A1	2011-12-01	Kurn	
	47	20110319290	A1	2011-12-29	Raymond et al.	
	48	20120003657	A1	2012-01-05	Myllykangas et al.	
	49	20120028310	A1	2012-02-02	Kurn et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	50	20120045797	A1	2012-02-23	Kurn et al.	
	51	20120071331	A1	2012-03-22	Casbon et al.	
	52	20120074925	A1	2012-03-29	Oliver	
	53	20120102054	A1	2012-04-26	POPESCU et al.	
	54	20120107811	A1	2012-05-03	Kelso et al.	
	55	20120122701	A1	2012-05-17	Ryan et al.	
	56	20120149068	A1	2012-06-14	Kurn	
	57	20120156728	A1	2012-06-21	Li et al.	
	58	20120157322	A1	2012-06-21	Myllykangas et al.	
	59	20120190587	A1	2012-07-26	Kurn et al.	
	60	20120208705	A1	2012-08-16	Steemers et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	61	20120220483	A1	2012-08-30	KURN et al.	
	62	20120220494	A1	2012-08-30	Samuels et al.	
	63	20120237943	A1	2012-09-20	Soldatov et al.	
	64	20120238738	A1	2012-09-20	Hendrickson	
	65	20120245041	A1	2012-09-27	Brenner et al.	
	66	20120270212	A1	2012-10-25	RABINOWITZ et al.	
	67	20120283145	A1	2012-11-08	Wang	
	68	20120289426	A1	2012-11-15	Roos et al.	
	69	20120309002	A1	2012-12-06	Link	
	70	20130005585	A1	2013-01-03	Anderson et al.	
	71	20130059738	A1	2013-03-07	LEAMON et al.	

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	72	20130231253	A1	2013-09-05	Amorese et al.	
	73	20140038188	A1	2014-02-06	Kurn	
	74	20140038236	A1	2014-02-06	Kurn et al.	
	75	20140065692	A1	2014-03-06	KURN et al.	
	76	20140274729	A1	2014-09-18	Kurn et al.	
	77	20140274731	A1	2014-09-18	RAYMOND et al.	
	78	20140274738	A1	2014-09-18	Amorese et al.	
	79	20140303000	A1	2014-10-09	Armour	
	80	20140378345	A1	2014-12-25	Hindson et al.	
	81	20150011396	A1	2015-01-08	Schroeder et al.	
	82	20150017635	A1	2015-01-15	Myllykangas et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	83	20150101595	A1	2015-04-16	Hancock et al.	
	84	20150133319	A1	2015-05-14	Fu et al.	
	85	20150284769	A1	2015-10-08	Schroeder	
	86	20150299767	A1	2015-10-22	Armour et al.	
	87	20150299784	A1	2015-10-22	Fan et al.	
	88	20150299812	A1	2015-10-22	Talasaz	
	89	20160122756	A1	2016-05-05	Armour	
	90	20160130576	A1	2016-05-12	Armour	
	91	20160153039	A1	2016-06-02	Amorese et al.	
	92	20160203259	A1	2016-07-14	SCOLNICK et al.	
	93	20160220994	A1	2016-08-04	Wright	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0430

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	94	20160251711	A1	2016-09-01	AMORESE et al.	
	95	20160265042	A1	2016-09-15	Schroeder et al.	
	96	20160275240	A1	2016-09-22	HUELGA et al.	
	97	20160296930	A1	2016-10-13	Matear 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]	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	2006/086668	WO	A2	2006-08-17	Epict Technologies		
	2	2006/137733	WO	A1	2006-12-28	Keygene NV		
	3	2007/018601	WO	A1	2007-02-15	Rubicon Genomics, Inc		
	4	2007/019444	WO	A2	2007-02-15	Euclid Diagnostics LLC		
	5	2007/030759	WO	A2	2007-03-15	Nugen Technologies, Inc		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	6	2007/037678	WO	A2	2007-04-05	Keygene NV		
	7	2007/052006	WO	A1	2007-05-10	Solexa Ltd		
	8	2007/057652	WO	A1	2007-05-24	Solexa Ltd		
	9	2007/073165	WO	A1	2007-06-28	Keygene NV		
	10	2007/136717	WO	A1	2007-11-29	Nugen Technologies, Inc		
	11	2008/005459	WO	A2	2008-01-10	Nugen Technologies, Inc		
	12	2008/015396	WO	A2	2008-02-07	Solexa Ltd		
	13	2008/033442	WO	A2	2008-03-20	Univ Leland Stanford Junior		
	14	2008/093098	WO	A2	2008-08-07	Solexa Ltd		
	15	2008/115185	WO	A2	2008-09-25	Nimblegen Systems, Inc		
	16	2008150432	WO	A1	2008-12-11	454 Life Sciences Corp		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	17	2009/053039	WO	A1	2009-04-30	Roche Diagnostics GmbH		
	18	2009/102878	WO	A2	2009-08-20	Nugen Technologies, Inc		
	19	2009/102896	WO	A2	2009-08-20	Nugen Technologies, Inc		
	20	2009/112844	WO	A1	2009-09-17	Stichting Het Nl Kanker I		
	21	2009/117698	WO	A2	2009-09-24	Nugen Technologies, Inc		
	22	2009/120372	WO	A2	2009-10-01	Pacific Biosciences Of California, Inc		
	23	2009/120374	WO	A2	2009-10-01	Pacific Biosciences Of California, Inc		
	24	2010/003153	WO	A2	2010-01-07	Life Technologies Corp		
	25	2010/030683	WO	A1	2010-03-18	Rosetta Inpharmatics LLC		
	26	2010/039991	WO	A2	2010-04-08	Texas A & M Univ Sys		
	27	2010/063711	WO	A1	2010-06-10	Angeletti P Ist Richerche Bio		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	28	2010/064893	WO	A1	2010-06-10	Keygene NV		
	29	2010/085715	WO	A1	2010-07-29	Quanta Biosciences		
	30	2010/115154	WO	A1	2010-10-07	Fluidigm Corp		
	31	2010/129937	WO	A2	2010-11-11	Life Technologies Corp		
	32	2010091246	WO	A2	2010-08-12	Univ Northwestern		
	33	2011/003630	WO	A1	2011-01-13	Max Planck Gesellschaft		
	34	2011/009941	WO	A1	2011-01-27	Olink Genomics AB		
	35	2011/019964	WO	A1	2011-02-17	Nugen Technologies, Inc		
	36	2011/032053	WO	A1	2011-03-17	Nugen Technologies, Inc		
	37	2011/053987	WO	A1	2011-05-05	Nugen Technologies, Inc		
	38	2011/151777	WO	A1	2011-12-08	Consiglio Nazionale Ricerche		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

IDT Exhibit 1019.0434

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	39	2011/156529	WO	A2	2011-12-15	Nugen Technologies, Inc		
	40	2011032040	WO	A1	2011-03-17	Centrillion Technology Holding Corp		
	41	2012/013932	WO	A1	2012-02-02	Univ St Andrews		
	42	2012/054873	WO	A3	2012-08-23	Cold Spring Harbor Lab		
	43	2012/061832	WO	A1	2012-05-10	Illumina, Inc		
	44	2012/103154	WO	A1	2012-08-02	Nugen Technologies, Inc		
	45	2013/059740	WO	A1	2013-04-25	Foundation Medicine, Inc		
	46	2013/059746	WO	A1	2013-04-25	Nugen Technologies, Inc		
	47	2013/112923	WO	A1	2013-08-01	Nugen Technologies, Inc		
	48	2013/130512	WO	A3	2013-10-24	Univ North Carolina		
	49	2013/138510	WO	A9	2014-07-10	Patel, Abhijit Ajit		

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /T.E.S/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16743604
	Filing Date	2020-01-15
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Attorney Docket Number	NUGN-001/03US 34079/516

	50	2013/177220	WO	A1	2013-11-28	Scripps Research Inst		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

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	Westin 2000, Anchored multiplex amplification on a microelectronic chip array, Nat Biotech 18:199-204	
	2	Wienholds, 2004, Target-selected gene inactivation in zebrafish, Meth Cell Biol 77:69-90	
	3	Wolford, 2000, High-throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC), Hum Genet 107:483-487	
	4	Xi, 2011, Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion, PNAS 108(46):e1128-e1136	
	5	Xu, 2012, FastUniq: A fast de novo duplicates removal tool for paired short reads, PLoSOne 7(12):e52249	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/TERESA E STRZELECKA/	Date Considered	12/02/2020

*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. /T.E.S/

Web Search History

date, time	web site	search string
12/2/2020 1:05:42 PM	Google	sequencing "duplicate read" removal barcode
12/2/2020 1:29:15 PM	Google	sequencing barcodesequenced separately
12/2/2020 1:29:22 PM	Google	sequencing barcode sequenced separately
12/2/2020 1:29:34 PM	Google	sequencing barcode "sequenced separately"
12/2/2020 1:48:34 PM	Google	sequencing "barcode sequenced separately"

IDT Exhibit 1019.0437

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 16/743,604	Filing Date 01/15/2020	☐ 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	03/05/2021	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 10	Minus	** 21	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 1	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".	/TERRY BRYANT/
*** 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.

Electronic Acknowledgement Receipt

EFS ID:	42097320
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	05-MAR-2021
Filing Date:	15-JAN-2020
Time Stamp:	10:49:21
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$170
RAM confirmation Number	E202135A49173164
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

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	33486 2b5885c950ff7a8084ef064f0403420f12fe3513	no	2
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30501 bc5196d8634b6e325e82c06c784ebbd415407e46	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	63987		

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.

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

Application No.: 16743604

Filing Date: 15-Jan-2020

Applicant/Patent under Reexamination: Amorese

Electronic Terminal Disclaimer filed on March 5, 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

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

Electronic Petition Request	**TERMINAL DISCLAIMER TO OBVIATE A DOUBLE PATENTING REJECTION OVER A "PRIOR" PATENT**
Application Number	16743604
Filing Date	15-Jan-2020
First Named Inventor	Douglas Amorese
Attorney Docket Number	NUGN-001/03US 34079/516
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

☒ 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
NUGEN TECHNOLOGIES, INC.	100%

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

9546399

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

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

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

○	I certify, in accordance with 37 CFR 1.4(d)(4), that the terminal disclaimer fee under 37 CFR 1.20(d) required for this terminal disclaimer has already been paid in the above-identified application.

Applicant claims the following fee status:

○ Small Entity

○ Micro Entity

◉ Regular Undiscounted

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

THIS PORTION MUST BE COMPLETED BY THE SIGNATORY OR SIGNATORIES

I certify, in accordance with 37 CFR 1.4(d)(4) that I am:

◉ An attorney or agent registered to practice before the Patent and Trademark Office who is of record in this application

Registration Number 36989

○ 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	/Thomas C. Meyers/
Name	Thomas C. Meyers

*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:	16743604
Filing Date:	15-Jan-2020
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Filer:	Thomas C. Meyers/Sandra Richards
Attorney Docket Number:	NUGN-001/03US 34079/516

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**

AMENDMENTS TO THE CLAIMS

Pursuant to 37 C.F.R. §1.121(c), this listing of claims will replace all prior versions and listings of claims in this application.

1. - 21. (Cancelled)

22. (Currently Amended) A method for detecting duplicate sequencing reads, the method comprising:

ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides, an indexing site unique to a subset of the adaptors, and a primer binding site;

amplifying the adaptor-ligated fragments into amplicons;

sequencing the amplicons to produce sequence reads that include identifier and target sequences; and

identifying sequence reads with identical identifier and target sequences as duplicates.

23. (Previously Presented) The method of claim 1, further comprising removing a duplicate read from the sequence reads.

24. (Previously Presented) The method of claim 1, wherein the identifier site is sequenced with the target sequence.

25. (Previously Presented) The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.

26. (Currently Amended) The method of claim 1, wherein the nucleic acid fragments [[are]] comprise RNA.

IDT Exhibit 1019.0446

27. (Currently Amended) The method of claim 1[[26]], wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

28. (Previously Presented) The method of claim 1, wherein the primer binding site is a universal target sequencing primer binding site.

29. (Previously Presented) The method of claim 1, wherein each unique identifier is 6 nucleotides in length.

30. (Previously Presented) The method of claim 1, wherein each unique identifier is 8 nucleotides in length.

31. (Previously Presented) The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.

32. (Cancelled).

IDT Exhibit 1019.0447

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	NuGEN Technologies, Inc.	ART UNIT:	1637
SERIAL NUMBER:	16/743,604	CONF. NO.:	3249
FILING DATE:	January 15, 2020	EXAMINER:	Teresa Strzelecka
TITLE:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		

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

FILED ELECTRONICALLY

RESPONSE TO NON-FINAL OFFICE ACTION

This Amendment is in response to the Non-Final Office Action mailed December 8, 2020 from the United States Patent and Trademark Office.

Amendments to the Claims begin on page 2.

Remarks begin on page 4.

Applicant believes that no fees are due in connection with this Amendment. However, Applicant authorizes any required fees to be charged to Deposit Account No. 500369.

Electronic Acknowledgement Receipt

EFS ID:	42097289
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	05-MAR-2021
Filing Date:	15-JAN-2020
Time Stamp:	10:46:41
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		NUGN-001-03US-Amendment-3-5-2021.pdf	38087 ecb2f5641678268463c9f62a16ad557ecfda7203	yes	8

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Amendment/Req. Reconsideration-After Non-Final Reject	1	1
	Claims	2	3
	Applicant Arguments/Remarks Made in an Amendment	4	8
Warnings:			
Information:			
	Total Files Size (in bytes):	38087	

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 22-32 were pending. By this amendment, claims 22 and 26-27 are amended and claim 32 cancelled. Claims 26 and 27 have been amended merely for clarity. Upon entry of this amendment, claims 22-31 will be pending, with claim 22 the sole independent claim.

Support for the amendments to claim 22 claims is found throughout the subject application as originally filed, including at paragraphs [0008], [0013], [0026], [0030]-[0031], [0045], and [0064]-[0070].

No new matter is added.

Rejections Under 35 U.S.C. §103

The Non-Final Office Action states on pages 3-4 that claims 22-31 are rejected as obvious over US 2012/0220494 to Samuels et al., in view of McCloskey et al., "Encoding PCR Products with Batch-stamps and Barcodes", *Biochem. Genet.*, vol. 45, pp. 761-767 (2007) in further view of Illumina publication No. 770-2008-011 (2008).

Applicant submits that amended, independent claim 1, and all claims dependent therefrom, are non-obvious over any combination Samuels, McCloskey and the Illumina publication.

A *prima facie* case of obviousness requires a finding that all elements of a claimed invention are found in a combination of prior art references. *Par Pharm., Inc. v. TWI Pharm., Inc.*, 773 F.3d 1186, 1194 (Fed. Cir. 2014). Establishing a *prima facie* case also requires that the prior art provide a motivation to modify a reference or to combine multiple references in a manner that would lead one of ordinary skill in the art to an applicant's claimed invention with a reasonable expectation of success. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398 (2007); and MPEP §2143. Further if a proposed modification or combination of prior art references would change the principle of operation of a reference being modified, then the teachings of the references are

IDT Exhibit 1019.0451

not sufficient to render the claims *prima facie* obvious. *In re Ratti,* 270 F.2d 810, 813 (CCPA 1959).

Amended claim 1 recites a method for removing duplicate sequence reads, for example, from a pooled multiplex sequencing run. The claims recite ligating adaptors that include an identifier portion and unique indexing sites to nucleic acid fragments prior to sequencing.

The cited references, alone or in combination, fail to teach the use of adaptors that include indexing sites and identifiers in pooled sequencing reactions.

As the Office Action acknowledges at page 5, Samuels does not teach detecting duplicate sequence reads. Rather, Samuels reports using barcodes to differentiate sequencing errors from single nucleotide polymorphisms. *Samuels, [0339]-[0341].* In the methods of Samuels, nucleic acid fragments have an attached barcode *Id.* The barcodes and nucleic acids are sequenced. *Id.* The sequence reads, which each include the barcode sequence and fragment sequence, are compared to identify sequences with the same barcode but different fragment sequences. *Id.* This can elucidate sequencing errors from faithfully sequenced single nucleotide polymorphisms. *Id.*

However, Samuels fails to disclose detecting duplicate sequence reads by any means, let alone using an adapter with both an identifier and index site.

McCloskey cannot cure the omissions of Samuels.

McCloskey reports methods to identify redundant sequences using target-specific "encoding oligonucleotides". *McCloskey, page 766.* The target-specific "encoding oligonucleotides" hybridize to nucleic acids in a sample and include a "batch-stamp", a "barcode", and a target-specific "genomic target-binding site". *Id.* The "batch-stamp" marks a fragment as from a particular date/sample. *Id.* The "barcode" is a series of 7 random bases that identify individual fragments. *Id.*

In McCloskey, the presence of redundant sequences is determined by identifying sequence reads that include the same "batch-stamp" and "barcode". However, in contrast to the claimed methods, McCloskey does not use the sequence of a nucleic acid fragment to identify

IDT Exhibit 1019.0452

redundant or duplicate sequences. Thus, like Samuels, McCloskey does not teach identifying sequence reads with identical identifier and fragment sequences as recited in the pending claims.

Furthermore, the claimed methods provide advantages over those disclosed in McCloskey. For example, McCloskey teaches barcodes consisting of 7 nucleotides, which allows for 16,384 unique combinations of adaptors. *Id., page 764.* However, next-generation sequencers have are able to sequence far more than 16,384 fragments during a single sequencing run. Thus, using the methods of McCloskey, in a single sample of 250,000 fragments attached to "encoding nucleotides" with 7 nucleotide "barcodes" there will be 15 *different* fragments that share the exact same "barcode" and "batch-stamp". Yet, using McCloskey, these sequences would be misidentified as duplicates.

This is obviated in the claimed methods by using the identifier and fragment sequence to identify duplicate reads. Moreover, because the fragment sequence is used in conjunction with the identifier to find duplicates, the identifier can be less complex than the "barcode" in McCloskey. The lower complexity of adaptors comprising fewer nucleotides proves more cost-effective, while maintaining the ability to correctly identify true duplications.

Furthermore, and in direct contrast to the claimed methods, McCloskey does not use adaptors. Rather, McCloskey uses oligonucleotides with a target-specific "genomic target-binding site". Thus, using McCloskey with samples containing a large number of fragments would provide prohibitively costly as a unique primer must be made for every single fragment sequenced.

Moreover, Samuels cannot be modified by McCloskey to arrive at the claimed methods, as such a modification would change the principle of operation of Samuels. As explained Samuels uses barcodes contrary to the identifier recited in the pending claims. That is, Samuels uses barcodes to identify sequences that are different and thus not duplicate. Modifying Samuels to use identifiers and fragment sequences to detect duplicate sequences would run counter to the principle of operation of Samuels, which seeks to *distinguish* between sequence reads with the

same barcode/identifier. As the principle of operation would be changed by this proposed modification, "the teachings of the reference are not sufficient to render the claims *prima facie* obvious". *MPEP 2143.01.*

The Illumina publication also does not remedy the deficiencies in Samuels and McCloskey.

The Illumina publication merely reports the use of adaptors with index sequences for multiplex sequencing. *Illumina, page 1*. The publication discloses that adaptors contain primer binding sites on both sides of the index sequence, which allows sequencing of the index sequence from one primer and sequencing of the fragment of interest from another primer. *Id., page 5.*

However, like Samuels and McCloskey, the Illumina publication does not teach or suggest adaptors having distinct identifiers and index sites and using the identifier and fragment sequence for deduplication.

Accordingly, based on the foregoing remarks, amended claim 22 and all claims dependent therefrom, are non-obvious over the cited references. Therefore, Applicant respectfully requests that the rejections under 35 U.S.C. §103 be reconsidered and withdrawn.

Double Patenting

The Office Action states on page 8 that claims 22, 23, and 26-31 are rejected on the ground of non-statutory double patenting over claims 1, 2, 9, 10, 13, 14, and 16 of U.S. Patent No. 9,546,399.

Applicant is filing herewith a terminal disclaimer over U.S. Patent No. 9,546,399. Accordingly, Applicant requests that the rejections on the ground of non-statutory double patenting be withdrawn.

IDT Exhibit 1019.0454

If there are any questions concerning this application, please contact Applicant's undersigned representative.

Dated: March 5, 2021

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: 617-856-8483

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	Art Unit
	TERESA E STRZELECKA	1637

CPC - Searched*		
Symbol	Date	Examiner
C12Q1/6869 |C12Q2525/191 | C12Q2535/122 plus text search	12/02/2020	/TS/
C12Q1/6869 |C12Q2525/191 | C12Q2535/122 plus text search	04/12/2021	/TS/

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
Inventor name search	12/02/2020	/TS/
WEST	12/02/2020	/TS/
Google	12/02/2020	/TS/
IDS (7/22/20)	12/02/2020	/TS/
Google	04/12/2021	/TS/
WEST	04/12/2021	/TS/

IDT Exhibit 1019.0456

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	Art Unit
	TERESA E STRZELECKA	1637

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner
	C12Q1/6869 |C12Q2525/191 | C12Q2535/122 plus text search	04/12/2021	/TS/

IDT Exhibit 1019.0457

Web Search History

date, time	web site	search string
4/12/2021 3:56:01 PM	Google	sequencing "duplicate reads" identification
4/12/2021 4:09:03 PM	Google	sequencing "duplicate reads" identificationbarcodes
4/12/2021 4:11:15 PM	Google	sequencing "duplicate reads" identification barcodes

IDT Exhibit 1019.0458

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	Art Unit
	TERESA E STRZELECKA	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	12/03/2020	04/12/2021							
	22	✓	=							
	23	✓	=							
	24	✓	=							
	25	✓	=							
	26	✓	=							
	27	✓	=							
	28	✓	=							
	29	✓	=							
	30	✓	=							
	31	✓	=							
	32	✓	-							

IDT Exhibit 1019.0459

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

21710 7590 04/21/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER
STRZELECKA, TERESA E

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 04/21/2021

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516	3249

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

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	07/21/2021

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)

21710 7590 04/21/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

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.
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516	3249

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

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	07/21/2021

EXAMINER	ART UNIT	CLASS-SUBCLASS
STRZELECKA, TERESA E	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. _____

IDT Exhibit 1019.0461

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.
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516	3249

21710 7590 04/21/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
STRZELECKA, TERESA E	
ART UNIT	**PAPER NUMBER**
1637	

DATE MAILED: 04/21/2021

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.

IDT Exhibit 1019.0462

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. 16/743,604	Applicant(s) Amorese et al.	
	Examiner TERESA E STRZELECKA	Art Unit 1637	AIA (FITF) Status Yes

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

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

1. ☑ This communication is responsive to amendment filed March 5, 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 22-31 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

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

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

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

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

Attachment(s)

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

/TERESA E STRZELECKA/
Primary Examiner, Art Unit 1637

IDT Exhibit 1019.0464

DETAILED ACTION

Notice of Pre-AIA or AIA Status

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

2. This office action is in response to an amendment filed March 5, 2021. Claims 22-32 were previously pending. Applicant amended claims 22, 26 and 27 and cancelled claim 32. Claims 22-31 are under consideration.

3. Applicant's claim amendments and arguments overcame the previously presented claim rejections. No new references were found teaching or suggesting the claims as amended.

Terminal Disclaimer

4. The terminal disclaimer filed on March 5, 2021 disclaiming the terminal portion of any patent granted on this application which would extend beyond the expiration date of U.S. Patent 9,946,399 has been reviewed and is accepted. The terminal disclaimer has been recorded.

In view of the acceptance of the terminal disclaimer, the previously presented obviousness-type double patenting rejection is withdrawn.

5. In conclusion, claims 22-31 are allowed.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to TERESA E STRZELECKA whose telephone number is (571)272-0789. The examiner can normally be reached on M-F 9-7.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is

encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, 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 https://ppair-my.uspto.gov/pair/PrivatePair. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/TERESA E STRZELECKA/
Primary Examiner, Art Unit 1637
April 12, 2021

WEST Search History for Application 16743604

Creation Date: 2021041218:29

Interference Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
sequencing and read near3 (duplicat $ 8 or redun $ 8) and (barcod $ 8 or index $ 8 or tag $ 8 or identifier) near3 (random or degenerat $ 8) and (duplicat $ 8 or redund $ 8) near3 remov $ 8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	PGPB, USPT	64	ADJ	YES	ASSIGNEE	04-12-2021

Prior Art Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
("20010000077" | "20010031739" | "20010034048" | "20010041334" | "20020028447" | "20020058270" | "20020115088" | "20020150919" | "20020155451" | "20020164628" | "20020164634" | "20020197639" | "20030017591" | "20030022207" | "20030082543" | "20030087251" | "20030119150" | "20030143555" | "20030175780" | "20030180779" | "20030186234" | "20030207279" | "20030211616" | "20030215926" | "20030224439" | "20030232348" | "20040002371" | "20040005614" | "20040023271" | "20040115815" | "20040137456" | "20040161742" | "20040203019" | "20040203025" | "20040248153" | "20050003441" | "20050014192" | "20050019793" | "20050059048" | "20050064456" | "20050123956" | "20050136417" | "20050142577" | "20050191656" | "20050191682" | "20050208538" | "20060008824" | "20060014182" | "20060024678" | "20060024711" | "20060035274" | "20060046251" | "20060051789" | "20060068415" | "20060134633" | "20060216724" | "20060263789" | "20060281082" |	PGPB, USPT	n/a	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0467

"20060286566"	"20060292597"						
"20070031857"	"20070134128"						
"20070141604"	"20070224607"						
"20070224613"	"20070231823"						
"20070238122"	"20070263045"						
"20080038727"	"20080087826"						
"20080103058"	"20080131937"						
"20080160580"	"20080176311"						
"20080182300"	"20080194413"						
"20080194416"	"20080206764"						
"20080213770"	"20080217246"						
"20080241831"	"20080242560"						
"20080286795"	"20090011959"						
"20090024331"	"20090026082"						
"20090029385"	"20090036663"						
"20090061425"	"20090061439"						
"20090068645"	"20090068655"						
"20090068709"	"20090105081"						
"20090117573"	"20090117621"						
"20090124514"	"20090127589"						
"20090130721"	"20090203085"						
"20090203531"	"20090233802"						
"20090233804"	"20090239232"						
"20090275486"	"20090280538"						
"20090298075"	"20100015666"						
"20100021973"	"20100022403"						
"20100029511"	"20100105052"						
"20100113296"	"20100129879"						
"20100137143"	"20100159559"						
"20100167954"	"20100173394"						
"20100203597"	"20100267043"						
"20100273219"	"20100311066"						
"20100323348"	"20110015096"						
"20110039732"	"20110104785"						
"20110105364"	"20110129827"						
"20110189679"	"20110224105"						
"20110288780"	"20110294132"						
"20110319290"	"20120003657"						
"20120028310"	"20120045797"						
"20120071331"	"20120074925"						
"20120102054"	"20120107811"						
"20120122701"	"20120149068"						
"20120156728"	"20120157322"						
"20120190587"	"20120208705"						
"20120220483"	"20120220494"						
"20120237943"	"20120238738"						
"20120245041"	"20120270212"						
"20120283145"	"20120289426"						
"20120309002"	"20130005585"						
"20130059738"	"20130231253"						
"20140038188"	"20140038236"						

IDT Exhibit 1019.0468

"20140065692" | "20140274729" |						
"20140274731" | "20140274738" |						
"20140303000" | "20140378345" |						
"20150011396" | "20150017635" |						
"20150101595" | "20150133319" |						
"20150284769" | "20150299767" |						
"20150299784" | "20150299812" |						
"20160122756" | "20160130576" |						
"20160153039" | "20160203259" |						
"20160220994" | "20160251711" |						
"20160265042" | "20160275240" |						
"20160296930" | "4362867" | "4458066" |						
"4469863" | "4582877" | "4876187" |						
"4925065" | "4935357" | "4942124" |						
"4988617" | "4996143" | "5011769" |						
"5034506" | "5035996" | "5043272" |						
"5082830" | "5090591" | "5130238" |						
"5169766" | "5171534" | "5194370" |						
"5216141" | "5234809" | "5235033" |						
"5242794" | "5384242" | "5386023" |						
"5399491" | "5409818" | "5418149" |						
"5422271" | "5427929" | "5480784" |						
"5494810" | "5508169" | "5508178" |						
"5510270" | "5525471" | "5545522" |						
"5554516" | "5554517" | "5556752" |						
"5565340" | "5573913" | "5578832" |						
"5589339" | "5602240" | "5637684" |						
"5641658" | "5644048" | "5665549" |						
"5667976" | "5667979" | "5679512" |						
"5681726" | "5683879" | "5688648" |						
"5705628" | "5708154" | "5710028" |						
"5712126" | "5716785" | "5726329" |						
"5750341" | "5759822" | "5763178" |						
"5789206" | "5824517" | "5824518" |						
"5837832" | "5876976" | "5882867" |						
"5888779" | "5888819" | "5945313" |						
"5952176" | "5958681" | "5965409" |						
"5969119" | "5972618" | "6004744" |						
"6004745" | "6027889" | "6027923" |						
"6030774" | "6037152" | "6056661" |						
"6077674" | "6087103" | "6090553" |						
"6090591" | "6107023" | "6110709" |						
"6150112" | "6159685" | "6160105" |						
"6169194" | "6172208" | "6174680" |						
"6180338" | "6190865" | "6194211" |						
"6197501" | "6197557" | "6210891" |						
"6225109" | "6225451" | "6232104" |						
"6251639" | "6262490" | "6270961" |						
"6280935" | "6287766" | "6287825" |						
"6291170" | "6306365" | "6306597" |						
"6309843" | "6326142" | "6335167" |						

IDT Exhibit 1019.0469

"6339147" | "6440705" | "6449562" | "6582938" | "6670461" | "6686156" | "6692918" | "6770748" | "6777180" | "6794499" | "6815164" | "6815167" | "6825011" | "6833246" | "6849404" | "6858413" | "6913884" | "6917726" | "6924104" | "6946251" | "7001724" | "7033764" | "7048481" | "7052847" | "7056676" | "7056716" | "7060441" | "7094536" | "7115400" | "7170050" | "7175982" | "7176025" | "7189512" | "7211390" | "7232656" | "7244559" | "7244567" | "7264929" | "7273730" | "7276720" | "7294461" | "7300755" | "7302146" | "7313308" | "7315019" | "7323305" | "7335762" | "7351557" | "7354717" | "7361466" | "7361468" | "7402386" | "7405281" | "7414117" | "7416844" | "7462452" | "7462468" | "7476503" | "7476504" | "7491498" | "7501245" | "7579153" | "7704687" | "7741463" | "7771934" | "7771946" | "7803550" | "7846666" | "7846733" | "7867703" | "7939258" | "7948015" | "7985565" | "8017335" | "8034568" | "8053192" | "8071311" | "8143001" | "8209130" | "8334116" | "8465950" | "8492095" | "8512956" | "8551709" | "8852867" | "8999677" | "9175325" | "9175336" | "9181582" | "9206418" | "9248076" | "9546399" | "9702004" | "9745627" | "9896721").PN.						
("4362867" | "4458066" | "4469863" | "4582877" | "4876187" | "4925065" | "4935357" | "4942124" | "4988617" | "4996143" | "5011769" | "5034506" | "5035996" | "5043272" | "5082830" | "5090591" | "5130238" | "5169766" | "5171534" | "5194370" | "5216141" | "5234809" | "5235033" | "5242794" | "5384242" | "5386023" | "5399491" | "5409818" | "5418149" | "5422271" | "5427929" | "5480784" | "5494810" | "5508169" | "5508178" | "5510270" | "5525471" | "5545522" | "5554516" | "5554517" | "5556752" | "5565340" | "5573913" | "5578832" | "5589339" | "5602240" | "5637684" | "5641658" | "5644048" | "5665549" | "5667976" | "5667979" | "5679512" | "5681726" |	USPT	214	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0470

"5683879" | "5688648" | "5705628" |						
"5708154" | "5710028" | "5712126" |						
"5716785" | "5726329" | "5750341" |						
"5759822" | "5763178" | "5789206" |						
"5824517" | "5824518" | "5837832" |						
"5876976" | "5882867" | "5888779" |						
"5888819" | "5945313" | "5952176" |						
"5958681" | "5965409" | "5969119" |						
"5972618" | "6004744" | "6004745" |						
"6027889" | "6027923" | "6030774" |						
"6037152" | "6056661" | "6077674" |						
"6087103" | "6090553" | "6090591" |						
"6107023" | "6110709" | "6150112" |						
"6159685" | "6160105" | "6169194" |						
"6172208" | "6174680" | "6180338" |						
"6190865" | "6194211" | "6197501" |						
"6197557" | "6210891" | "6225109" |						
"6225451" | "6232104" | "6251639" |						
"6262490" | "6270961" | "6280935" |						
"6287766" | "6287825" | "6291170" |						
"6306365" | "6306597" | "6309843" |						
"6326142" | "6335167" | "6339147" |						
"6440705" | "6449562" | "6582938" |						
"6670461" | "6686156" | "6692918" |						
"6770748" | "6777180" | "6794499" |						
"6815164" | "6815167" | "6825011" |						
"6833246" | "6849404" | "6858413" |						
"6913884" | "6917726" | "6924104" |						
"6946251" | "7001724" | "7033764" |						
"7048481" | "7052847" | "7056676" |						
"7056716" | "7060441" | "7094536" |						
"7115400" | "7170050" | "7175982" |						
"7176025" | "7189512" | "7211390" |						
"7232656" | "7244559" | "7244567" |						
"7264929" | "7273730" | "7276720" |						
"7294461" | "7300755" | "7302146" |						
"7313308" | "7315019" | "7323305" |						
"7335762" | "7351557" | "7354717" |						
"7361466" | "7361468" | "7402386" |						
"7405281" | "7414117" | "7416844" |						
"7462452" | "7462468" | "7476503" |						
"7476504" | "7491498" | "7501245" |						
"7579153" | "7704687" | "7741463" |						
"7771934" | "7771946" | "7803550" |						
"7846666" | "7846733" | "7867703" |						
"7939258" | "7948015" | "7985565" |						
"8017335" | "8034568" | "8053192" |						
"8071311" | "8143001" | "8209130" |						
"8334116" | "8465950" | "8492095" |						
"8512956" | "8551709" | "8852867" |						
"8999677" | "9175325" | "9175336" |						

"9181582" | "9206418" | "9248076" | "9546399" | "9702004" | "9745627" | "9896721").PN.						
(("4362867" | "4458066" | "4469863" | "4582877" | "4876187" | "4925065" | "4935357" | "4942124" | "4988617" | "4996143" | "5011769" | "5034506" | "5035996" | "5043272" | "5082830" | "5090591" | "5130238" | "5169766" | "5171534" | "5194370" | "5216141" | "5234809" | "5235033" | "5242794" | "5384242" | "5386023" | "5399491" | "5409818" | "5418149" | "5422271" | "5427929" | "5480784" | "5494810" | "5508169" | "5508178" | "5510270" | "5525471" | "5545522" | "5554516" | "5554517" | "5556752" | "5565340" | "5573913" | "5578832" | "5589339" | "5602240" | "5637684" | "5641658" | "5644048" | "5665549" | "5667976" | "5667979" | "5679512" | "5681726" | "5683879" | "5688648" | "5705628" | "5708154" | "5710028" | "5712126" | "5716785" | "5726329" | "5750341" | "5759822" | "5763178" | "5789206" | "5824517" | "5824518" | "5837832" | "5876976" | "5882867" | "5888779" | "5888819" | "5945313" | "5952176" | "5958681" | "5965409" | "5969119" | "5972618" | "6004744" | "6004745" | "6027889" | "6027923" | "6030774" | "6037152" | "6056661" | "6077674" | "6087103" | "6090553" | "6090591" | "6107023" | "6110709" | "6150112" | "6159685" | "6160105" | "6169194" | "6172208" | "6174680" | "6180338" | "6190865" | "6194211" | "6197501" | "6197557" | "6210891" | "6225109" | "6225451" | "6232104" | "6251639" | "6262490" | "6270961" | "6280935" | "6287766" | "6287825" | "6291170" | "6306365" | "6306597" | "6309843" | "6326142" | "6335167" | "6339147" | "6440705" | "6449562" | "6582938" | "6670461" | "6686156" | "6692918" | "6770748" | "6777180" | "6794499" | "6815164" | "6815167" | "6825011" | "6833246" | "6849404" | "6858413" | "6913884" | "6917726" | "6924104" | "6946251" | "7001724" | "7033764" |	USPT	2	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0472

"7048481" | "7052847" | "7056676" | "7056716" | "7060441" | "7094536" | "7115400" | "7170050" | "7175982" | "7176025" | "7189512" | "7211390" | "7232656" | "7244559" | "7244567" | "7264929" | "7273730" | "7276720" | "7294461" | "7300755" | "7302146" | "7313308" | "7315019" | "7323305" | "7335762" | "7351557" | "7354717" | "7361466" | "7361468" | "7402386" | "7405281" | "7414117" | "7416844" | "7462452" | "7462468" | "7476503" | "7476504" | "7491498" | "7501245" | "7579153" | "7704687" | "7741463" | "7771934" | "7771946" | "7803550" | "7846666" | "7846733" | "7867703" | "7939258" | "7948015" | "7985565" | "8017335" | "8034568" | "8053192" | "8071311" | "8143001" | "8209130" | "8334116" | "8465950" | "8492095" | "8512956" | "8551709" | "8852867" | "8999677" | "9175325" | "9175336" | "9181582" | "9206418" | "9248076" | "9546399" | "9702004" | "9745627" | "9896721").PN.) and sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8)						
("20010000077" | "20010031739" | "20010034048" | "20010041334" | "20020028447" | "20020058270" | "20020115088" | "20020150919" | "20020155451" | "20020164628" | "20020164634" | "20020197639" | "20030017591" | "20030022207" | "20030082543" | "20030087251" | "20030119150" | "20030143555" | "20030175780" | "20030180779" | "20030186234" | "20030207279" | "20030211616" | "20030215926" | "20030224439" | "20030232348" | "20040002371" | "20040005614" | "20040023271" | "20040115815" | "20040137456" | "20040161742" | "20040203019" | "20040203025" | "20040248153" | "20050003441" | "20050014192" | "20050019793" | "20050059048" | "20050064456" | "20050123956" | "20050136417" | "20050142577" | "20050191656" |	PGPB	181	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0473

"20050191682" | "20050208538" |							
"20060008824" | "20060014182" |							
"20060024678" | "20060024711" |							
"20060035274" | "20060046251" |							
"20060051789" | "20060068415" |							
"20060134633" | "20060216724" |							
"20060263789" | "20060281082" |							
"20060286566" | "20060292597" |							
"20070031857" | "20070134128" |							
"20070141604" | "20070224607" |							
"20070224613" | "20070231823" |							
"20070238122" | "20070263045" |							
"20080038727" | "20080087826" |							
"20080103058" | "20080131937" |							
"20080160580" | "20080176311" |							
"20080182300" | "20080194413" |							
"20080194416" | "20080206764" |							
"20080213770" | "20080217246" |							
"20080241831" | "20080242560" |							
"20080286795" | "20090011959" |							
"20090024331" | "20090026082" |							
"20090029385" | "20090036663" |							
"20090061425" | "20090061439" |							
"20090068645" | "20090068655" |							
"20090068709" | "20090105081" |							
"20090117573" | "20090117621" |							
"20090124514" | "20090127589" |							
"20090130721" | "20090203085" |							
"20090203531" | "20090233802" |							
"20090233804" | "20090239232" |							
"20090275486" | "20090280538" |							
"20090298075" | "20100015666" |							
"20100021973" | "20100022403" |							
"20100029511" | "20100105052" |							
"20100113296" | "20100129879" |							
"20100137143" | "20100159559" |							
"20100167954" | "20100173394" |							
"20100203597" | "20100267043" |							
"20100273219" | "20100311066" |							
"20100323348" | "20110015096" |							
"20110039732" | "20110104785" |							
"20110105364" | "20110129827" |							
"20110189679" | "20110224105" |							
"20110288780" | "20110294132" |							
"20110319290" | "20120003657" |							
"20120028310" | "20120045797" |							
"20120071331" | "20120074925" |							
"20120102054" | "20120107811" |							
"20120122701" | "20120149068" |							
"20120156728" | "20120157322" |							
"20120190587" | "20120208705" |							

IDT Exhibit 1019.0474

"20120220483" | "20120220494" | "20120237943" | "20120238738" | "20120245041" | "20120270212" | "20120283145" | "20120289426" | "20120309002" | "20130005585" | "20130059738" | "20130231253" | "20140038188" | "20140038236" | "20140065692" | "20140274729" | "20140274731" | "20140274738" | "20140303000" | "20140378345" | "20150011396" | "20150017635" | "20150101595" | "20150133319" | "20150284769" | "20150299767" | "20150299784" | "20150299812" | "20160122756" | "20160130576" | "20160153039" | "20160203259" | "20160220994" | "20160251711" | "20160265042").PN.						
(("20010000077" | "20010031739" | "20010034048" | "20010041334" | "20020028447" | "20020058270" | "20020115088" | "20020150919" | "20020155451" | "20020164628" | "20020164634" | "20020197639" | "20030017591" | "20030022207" | "20030082543" | "20030087251" | "20030119150" | "20030143555" | "20030175780" | "20030180779" | "20030186234" | "20030207279" | "20030211616" | "20030215926" | "20030224439" | "20030232348" | "20040002371" | "20040005614" | "20040023271" | "20040115815" | "20040137456" | "20040161742" | "20040203019" | "20040203025" | "20040248153" | "20050003441" | "20050014192" | "20050019793" | "20050059048" | "20050064456" | "20050123956" | "20050136417" | "20050142577" | "20050191656" | "20050191682" | "20050208538" | "20060008824" | "20060014182" | "20060024678" | "20060024711" | "20060035274" | "20060046251" | "20060051789" | "20060068415" | "20060134633" | "20060216724" | "20060263789" | "20060281082" | "20060286566" | "20060292597" | "20070031857" | "20070134128" | "20070141604" | "20070224607" |	PGPB	5	ADJ	YES	ASSIGNEE	12-02-2020

IDT Exhibit 1019.0475

"20070224613"	"20070231823"						
"20070238122"	"20070263045"						
"20080038727"	"20080087826"						
"20080103058"	"20080131937"						
"20080160580"	"20080176311"						
"20080182300"	"20080194413"						
"20080194416"	"20080206764"						
"20080213770"	"20080217246"						
"20080241831"	"20080242560"						
"20080286795"	"20090011959"						
"20090024331"	"20090026082"						
"20090029385"	"20090036663"						
"20090061425"	"20090061439"						
"20090068645"	"20090068655"						
"20090068709"	"20090105081"						
"20090117573"	"20090117621"						
"20090124514"	"20090127589"						
"20090130721"	"20090203085"						
"20090203531"	"20090233802"						
"20090233804"	"20090239232"						
"20090275486"	"20090280538"						
"20090298075"	"20100015666"						
"20100021973"	"20100022403"						
"20100029511"	"20100105052"						
"20100113296"	"20100129879"						
"20100137143"	"20100159559"						
"20100167954"	"20100173394"						
"20100203597"	"20100267043"						
"20100273219"	"20100311066"						
"20100323348"	"20110015096"						
"20110039732"	"20110104785"						
"20110105364"	"20110129827"						
"20110189679"	"20110224105"						
"20110288780"	"20110294132"						
"20110319290"	"20120003657"						
"20120028310"	"20120045797"						
"20120071331"	"20120074925"						
"20120102054"	"20120107811"						
"20120122701"	"20120149068"						
"20120156728"	"20120157322"						
"20120190587"	"20120208705"						
"20120220483"	"20120220494"						
"20120237943"	"20120238738"						
"20120245041"	"20120270212"						
"20120283145"	"20120289426"						
"20120309002"	"20130005585"						
"20130059738"	"20130231253"						
"20140038188"	"20140038236"						
"20140065692"	"20140274729"						
"20140274731"	"20140274738"						
"20140303000"	"20140378345"						

IDT Exhibit 1019.0476

"20150011396" | "20150017635" | "20150101595" | "20150133319" | "20150284769" | "20150299767" | "20150299784" | "20150299812" | "20160122756" | "20160130576" | "20160153039" | "20160203259" | "20160220994" | "20160251711" | "20160265042").PN.) and sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8)						
16/743604.app.	PGPB	1	ADJ	YES	ASSIGNEE	12-02-2020
sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8) and (duplicat$5 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	PGPB	47	ADJ	YES	ASSIGNEE	12-02-2020
sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8) and (duplicat$5 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	USPT	17	ADJ	YES	ASSIGNEE	12-02-2020
sequencing and read near3 (duplicat$5 or redund$8) and (barcod$5 or index$5 or identifier or tag$6) near3 (random or degener$8) and (duplicat$5 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	DWPI	2	ADJ	YES	ASSIGNEE	12-02-2020
16/743604.app.	PGPB	1	ADJ	YES	ASSIGNEE	04-12-2021
sequencing and read near3 (duplicat$8 or redun$8) and (barcod$8 or index$8 or tag$8 or identifier) near3 (random or degenerat$8) and (duplicat$8 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	PGPB	45	ADJ	YES	ASSIGNEE	04-12-2021
sequencing and read near3 (duplicat$8 or redun$8) and (barcod$8 or index$8 or tag$8 or identifier) near3 (random or degenerat$8) and (duplicat$8 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)	USPT	19	ADJ	YES	ASSIGNEE	04-12-2021
sequencing and read near3 (duplicat$8 or redun$8) and (barcod$8 or index$8 or tag$8	DWPI	2	ADJ	YES	ASSIGNEE	04-12-2021

IDT Exhibit 1019.0477

or identifier) near3 (random or degenerat$8) and (duplicat$8 or redund$8) near3 remov$8 and ((C12Q1/6869 | C12Q2525/191 | C12Q2535/122).CPC.)						

IDT Exhibit 1019.0478

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	1637

CPC						
Symbol					**Type**	**Version**
C12Q	/	1	/	6874	F	2013-01-01
C12Q	/	1	/	6869	I	2013-01-01
C12Q	/	2525	/	179	A	2013-01-01
C12Q	/	2525	/	191	A	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	/	191	I	1	3	2013-01-01
C12Q	/	2537	/	157	I	1	4	2013-01-01
C12Q	/	2563	/	179	I	1	5	2013-01-01
C12Q	/	2565	/	514	I	1	6	2013-01-01
C12Q	/	1	/	6869	I	2	1	2013-01-01
C12Q	/	2525	/	155	I	2	2	2013-01-01
C12Q	/	2525	/	179	I	2	3	2013-01-01
C12Q	/	2525	/	191	I	2	4	2013-01-01
C12Q	/	2535	/	122	I	2	5	2013-01-01
C12Q	/	2537	/	159	I	2	6	2013-01-01
C12Q	/	2563	/	179	I	2	7	2013-01-01

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 10	
/TERESA E STRZELECKA/ Primary Examiner, Art Unit 1637 (Primary Examiner)	12 April 2021 (Date)	O.G. Print Claim(s) 22	O.G. Print Figure None

U.S. Patent and Trademark Office

Part of Paper No.: 20210412

IDT Exhibit 1019.0479

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	1637

INTERNATIONAL CLASSIFICATION					
CLAIMED					
C12Q1/6874	/	1	/	6874	
C12Q1/6869	/	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:** 10	
/TERESA E STRZELECKA/ Primary Examiner, Art Unit 1637 (Primary Examiner)	12 April 2021 (Date)	O.G. Print Claim(s) 22	O.G. Print Figure None

IDT Exhibit 1019.0480

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	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 (Assistant Examiner)	(Date)	**Total Claims Allowed:** 10	
/TERESA E STRZELECKA/ Primary Examiner, Art Unit 1637 (Primary Examiner)	12 April 2021 (Date)	O.G. Print Claim(s) 22	O.G. Print Figure None

U.S. Patent and Trademark Office

Part of Paper No.: 20210412

IDT Exhibit 1019.0481

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.
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US ~~34079/516~~	3249

7590 04/30/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
STRZELECKA, TERESA E	
ART UNIT	**PAPER NUMBER**
1637	
NOTIFICATION DATE	**DELIVERY MODE**
04/30/2021	ELECTRONIC

NOTICE REQUIRING INVENTOR'S OATH OR DECLARATION

☐ An inventor's oath or declaration in compliance with 37 CFR 1.63 or substitute statement in compliance with 37 CFR 1.64 executed by or with respect to each inventor has not yet been submitted.

The inventor's oath or declaration or substitute statement dated 01/15/2020 is defective. The inventor's oath or declaration does not comply with 37 CFR 1.63 or the substitute statement does not comply with 37 CFR 1.64 because it:

☐ does not include a statement that the person executing the oath or declaration or substitute statement believes the named inventor or joint inventor to be the original inventor or an original joint inventor of a claimed invention in the application for which the oath or declaration is being submitted.

☐ does not state that the application was made or was authorized to be made by the person executing the oath or declaration or substitute statement.

☐ does not acknowledge that any willful false statement made in such oath or declaration or substitute statement is punishable under section 1001 of title 18 by fine or imprisonment of not more than 5 years, or both.

☑ does not identify the person executing the oath or declaration or substitute statement by his or her legal name.

☐ does not identify the application to which the oath or declaration or substitute statement is directed.

☑ Other: First name of 3rd inventor does not match ADS ("Ben SCHROEDER" vs. "Benjamin SCHROEDER")

The substitute statement dated _____ does not comply with 37 CFR 1.64 because it:

☐ does not identify the inventor with respect to whom the statement applies.

☐ does not identify the relationship to the non-signing inventor.

☐ does not identify the circumstances permitting execution of the statement.

☐ was signed by someone other than the applicant.

☐ Other: _____

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 inventor MUST be filed **no later than the date on which the issue fee is paid.** See 35 U.S.C. 115(f). Failure to timely comply will result in ABANDONMENT of this application. This period for reply is not extendable under 37 CFR 1.136(a).

Questions relating to this Notice should be directed to the Application Assistance Unit at 571-272-4200.

(571)-272-4200 or 1(888)-786-0101
Patent Publication Branch
Office of Data Management

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.
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516	3249

21710 7590 05/07/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

EXAMINER	
STRZELECKA, TERESA E	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
05/07/2021	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):

ip@brownrudnick.com
usactions@brownrudnick.com

IDT Exhibit 1019.0483

Notice of Allowability	Application No. 16/743,604	Applicant(s) Amorese et al.	
	Examiner TERESA E STRZELECKA	Art Unit 1637	AIA (FITF) Status Yes

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

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

1. ☑ This communication is responsive to Printer rush dated.
   ☐ 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 22-31 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

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

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

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

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

Attachment(s)

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

/TERESA E STRZELECKA/
Primary Examiner, Art Unit 1637

IDT Exhibit 1019.0484

DETAILED ACTION

Notice of Pre-AIA or AIA Status

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

2. This Corrected Notice of Allowability is issued to correct claim dependencies of claims 23-31, which should depend from claim 22.

EXAMINER'S AMENDMENT

3. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in an interview with Thomas Meyers on May 4, 2021.

The application has been amended as follows:

- In claim 23, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.
- In claim 24, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.
- In claim 25, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.
- In claim 26, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.
- In claim 27, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.
- In claim 28, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.
- In claim 29, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.
- In claim 30, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.

- In claim 31, line 1, delete ---1--- after "claim" and substitute ---22--- therefor.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to TERESA E STRZELECKA whose telephone number is (571)272-0789. The examiner can normally be reached on M-F 9-7.

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

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, 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 https://ppair-my.uspto.gov/pair/PrivatePair. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/TERESA E STRZELECKA/
Primary Examiner, Art Unit 1637
May 4, 2021

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	Art Unit
	TERESA E STRZELECKA	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	12/03/2020	04/12/2021							
	22	✓	=							
	23	✓	=							
	24	✓	=							
	25	✓	=							
	26	✓	=							
	27	✓	=							
	28	✓	=							
	29	✓	=							
	30	✓	=							
	31	✓	=							
	32	✓	-							

IDT Exhibit 1019.0487

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	1637

CPC						
Symbol					**Type**	**Version**
C12Q	/	1	/	6874	F	2013-01-01
C12Q	/	1	/	6869	I	2013-01-01
C12Q	/	2525	/	179	A	2013-01-01
C12Q	/	2525	/	191	A	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	/	191	I	1	3	2013-01-01
C12Q	/	2537	/	157	I	1	4	2013-01-01
C12Q	/	2563	/	179	I	1	5	2013-01-01
C12Q	/	2565	/	514	I	1	6	2013-01-01
C12Q	/	1	/	6869	I	2	1	2013-01-01
C12Q	/	2525	/	155	I	2	2	2013-01-01
C12Q	/	2525	/	179	I	2	3	2013-01-01
C12Q	/	2525	/	191	I	2	4	2013-01-01
C12Q	/	2535	/	122	I	2	5	2013-01-01
C12Q	/	2537	/	159	I	2	6	2013-01-01
C12Q	/	2563	/	179	I	2	7	2013-01-01

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	10	
/TERESA E STRZELECKA/ Primary Examiner, Art Unit 1637	04 May 2021	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	22	None

U.S. Patent and Trademark Office — Part of Paper No.: 20210504

IDT Exhibit 1019.0488

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	1637

INTERNATIONAL CLASSIFICATION				
CLAIMED				
C12Q1/6874	/	1	/	6874
C12Q1/6869	/	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:** 10	
/TERESA E STRZELECKA/ Primary Examiner, Art Unit 1637 (Primary Examiner) 04 May 2021 (Date)	O.G. Print Claim(s) 22	O.G. Print Figure None

U.S. Patent and Trademark Office Part of Paper No.: 20210504

IDT Exhibit 1019.0489

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/743,604	Amorese et al.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	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 (Assistant Examiner)	(Date)	**Total Claims Allowed:** 10	
/TERESA E STRZELECKA/ Primary Examiner, Art Unit 1637 (Primary Examiner)	04 May 2021 (Date)	O.G. Print Claim(s) 22	O.G. Print Figure None

U.S. Patent and Trademark Office

Part of Paper No.: 20210504

IDT Exhibit 1019.0490

Electronic Acknowledgement Receipt

EFS ID:	42977446
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Michelle Aiello
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	14-JUN-2021
Filing Date:	15-JAN-2020
Time Stamp:	14:52:22
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$140
RAM confirmation Number	E20216DE52361695
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Request under Rule 48 correcting inventorship	NUGN-001-03US-Request_for_Inventor_Related_Correction_PTO-AIA-40.pdf	210313 491d7baa993ea4581d05ad2d859b6de3ae63aad6	no	2
Warnings:					
Information:					
2	Application Data Sheet	NUGN-001-03US-Corrected_ADS.pdf	185920 27ea81452077779a5b20ba7af228beca9e4d76b5	no	5
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
3	Fee Worksheet (SB06)	fee-info.pdf	30643 ce5326414ce882d5918d0b2aa7a5992a406487ae	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	426876		

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.

Doc Code: R48.REQ
Document Description: Request under Rule 48 correcting inventorship

PTO/AIA/40 (04-18)
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 displays a valid OMB control number.

REQUEST FOR CORRECTION IN A PATENT APPLICATION RELATING TO INVENTORSHIP OR AN INVENTOR NAME, OR ORDER OF NAMES, OTHER THAN IN A REISSUE APPLICATION (37 CFR 1.48)		
	Application Number	16/743,604-Conf. #3249
	Filing Date	January 15, 2020
	First Named Inventor	Douglas A. Amorese
	Art Unit	1637
	Examiner Name	T. E. Strzelecka
	Practitioner Docket Number	NUGN-001/03US 34079/516

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

Applicant hereby **requests that the inventorship be corrected or changed, or that the name of the inventor or a joint inventor, or the order of the names of joint inventors, be changed,** in the above-identified application. **Note**: 37 CFR 1.48 applies to any request to correct inventorship filed on or after September 16, 2012, regardless of the application filing date. **Do not submit this form after payment of the issue fee or if the application has been patented.** See 37 CFR 1.324 for correction of inventorship in a patent.

Please check the applicable box(es) below.

For a nonprovisional application:

☐ 1.This request is to correct or change the inventorship in a **nonprovisional** application (**under 37 CFR 1.48(a)**) and includes:

☐ An application data sheet (ADS) in accordance with 37 CFR 1.76(c) with the corrected or updated information shown with markings (*e.g.*, underlining for insertions, strikethrough for deletions). See the Manual of Patent Examining Procedure (MPEP) section 601.05(a) for information about filing an ADS in an application filed on/after September 16, 2012. For information about filing a Supplemental ADS in an application filed before September 16, 2012, see MPEP 601.05(b).

☐ The processing fee set forth in 37 CFR 1.17(i). $ _____

☐ An inventor is being added. An inventor's oath or declaration by any actual inventor who has not yet executed an oath or declaration is required (see 37 CFR 1.48(b)). See MPEP 602.01(a) for information about an inventor's oath or declaration for an application filed on/after September 16, 2012 (*e.g.*, form **PTO/AIA/01**). For information about an inventor's oath or declaration for an application filed before September 16, 2012 (*e.g.*, form **PTO/SB/01**), see MPEP 602.01(b).

☐ This request is being filed after the first Office action on the merits has been given or mailed (see 37 CFR 1.48(c) and 1.17(d)). Check one of the following:

☐ This request to correct or change the inventorship is due solely to the cancellation of claims in the application.

OR

☐ The fee set forth in 37 CFR 1.17(d) is due (in addition to the fee set forth in 37 CFR 1.17(i)). $ _____

IDT Exhibit 1019.0493

PTO/AIA/40 (04-18)
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 displays a valid OMB control number.

REQUEST FOR CORRECTION IN A PATENT APPLICATION RELATING TO INVENTORSHIP OR AN INVENTOR NAME, OR ORDER OF NAMES, OTHER THAN IN A REISSUE APPLICATION (37 CFR 1.48)

[X] 2. This request is to correct or update the name of the inventor or a joint inventor, or the order of names of joint inventors, in a **nonprovisional** application (**under 37 CFR 1.48(f)**) and includes:

[X] An application data sheet in accordance with 37 CFR 1.76(c) identifying the complete inventive entity, including the corrected or updated name of the inventor, or the new order of names shown with markings (*e.g.*, underlining for insertions, strikethrough for deletions). See the MPEP 601.05(a) for information about filing an ADS in an application filed on/after September 16, 2012. For information about filing a Supplemental ADS in an application filed before September 16, 2012, see MPEP 601.05(b).

[X] The processing fee set forth in 37 CFR 1.17(i). $ 140.00

For a provisional application:

[ ] This request is to change or correct the inventorship, or correct or update the name of the inventor or a joint inventor, in a **provisional** application (**under 37 CFR 1.48(d)**) and includes:

[ ] **Attached hereto** is a document that is signed by a party set forth in 37 CFR 1.33(b) and identifies each inventor by his or her legal name, in the preferred order. Note: the document may be an application data sheet in accordance with 37 CFR 1.76(c) that identifies the changes with markings (underlining for insertions, strikethrough for deletions).

[ ] The processing fee set forth in 37 CFR 1.17(q). $ _____

Fee Payment Information:

[ ] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29.
Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously

[ ] A check in the amount of the fee is enclosed.

[ ] Payment by credit card. Form PTO-2038 is attached.

[X] The Director is hereby authorized to charge any fees which may be required, or credit any overpayment to Deposit Account No. 50-0369 .

[X] Payment made via EFS-Web.

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I am the

[ ] Applicant* [X] attorney or agent of record Registration number 36,989 [ ] attorney or agent acting under 37 CFR 1.34 Registration number _____

Signature /Thomas C. Meyers/

Typed or printed name Thomas C. Meyers

Date June 14, 2021

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. *Juristic entities must be represented by a patent practitioner (See 37 CFR 1.31, applicable to any paper filed on or after September 16, 2012 that is presented on behalf of a juristic entity, regardless of application filing date). Submit multiple forms if more than one signature is required, see below**.

[ ] ** Total of 1 forms are submitted.

64089253 v1

IDT Exhibit 1019.0494

Corrected Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	Douglas
Middle Name::	A.
Family Name::	Amorese
City of Residence::	Los Altos
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	900 Chesapeake Drive
City of mailing address::	Redwood City
State or Province of mailing address::	CA
Country of mailing address::	US
Postal or Zip Code of mailing address::	94063

Inventor Number::	2
Given Name::	Jonathan
Family Name::	Scolnick
City of Residence::	San Francisco
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	451 Kansas Street #518
City of mailing address::	San Francisco

State or Province of mailing address::	CA
Country of mailing address::	US
Postal or Zip Code of mailing address::	94107
Inventor Number::	3
Given Name::	~~Benjamin~~ Ben
Middle Name::	~~G.~~
Family Name::	Schroeder
City of Residence::	San Mateo
State or Province of Residence::	CA
Country of Residence::	US
Street of mailing address::	33 Hayward Ave, Apt 112
City of mailing address::	San Mateo
State or Province of mailing address::	CA
Country of mailing address::	US
Postal or Zip Code of mailing address::	94401

IDT Exhibit 1019.0496

Correspondence Information

Correspondence Customer Number::	21710
E-Mail address::	ip@brownrudnick.com

Application Information

Application Number::	16/743,604
Filing Date::	01/15/20
Application Type::	Nonprovisional
Subject Matter::	Utility
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
Attorney Docket Number::	NUGN-001/03US 34079/516
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets::	6
Small Entity?::	No
Petition included?::	No
Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy	No

IDT Exhibit 1019.0497

Order pursuant to 37 CFR 5.2::

Representative Information

Representative Customer Number:: 21710

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::	Prior Appl Status::	Parent Patent No.	Parent Issue Date
This Application	Continuation of	15/406002	01/13/17	~~Pending~~ Issued	10/570,448	02/25/2020
15/406002	Division of	14/540917	11/13/14	~~Pending~~ Issued	9,546,399	01/17/2017
14/540917	Claims benefit of provisional	61/903826	11/13/13	Expired	N/A	N/A

Foreign Priority Information

Applicant Information

Applicant Number:: 1

Applicant Type:: Assignee

Organization Name:: NUGEN TECHNOLOGIES, INC.

Street of mailing address:: 900 Chesapeake Drive

City of mailing address:: Redwood City

State or Province of mailing address:: CA

Country of mailing address:: US

Postal or Zip Code of mailing address:: 94063

IDT Exhibit 1019.0498

Assignee Information Including Non-Applicant Assignee Information

Signature:

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

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

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

Signature	/Thomas C. Meyers/	Date (YYYY-MM-DD)	2021-06-14
Name	Thomas C. Meyers	Registration Number	36,989

64089252 v1

IDT Exhibit 1019.0499

Electronic Patent Application Fee Transmittal

Application Number:	16743604
Filing Date:	15-Jan-2020
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Filer:	Thomas C. Meyers/Michelle Aiello
Attorney Docket Number:	NUGN-001/03US 34079/516

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:				
PROCESSING FEE, EXCEPT PROV. APPLS.	1830	1	140	140
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**140**

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
16/743,604	01/15/2020	1637	1820	NUGN-001/03US 34079/516	11	1

CONFIRMATION NO. 3249

UPDATED FILING RECEIPT

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

OC000000128344653

Date Mailed: 06/16/2021

Receipt is acknowledged of this non-provisional utility 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 FIRST INVENTOR, 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 corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction 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 provided that the request is grantable.

Inventor(s)

Douglas A. Amorese, Los Altos, CA;
Jonathan Scolnick, San Francisco, CA;
Ben Schroeder, San Mateo, CA;

Applicant(s)

NUGEN TECHNOLOGIES, INC., Redwood City, CA;

Power of Attorney: The patent practitioners associated with Customer Number 21710

Domestic Priority data as claimed by applicant

This application is a CON of 15/406,002 01/13/2017 PAT 10570448
which is a DIV of 14/540,917 11/13/2014 PAT 9546399
which claims benefit of 61/903,826 11/13/2013

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

IDT Exhibit 1019.0502

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

Projected Publication Date: Not Applicable

Non-Publication Request: No

Early Publication Request: No

Title

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

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).

IDT Exhibit 1019.0503

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
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516

CONFIRMATION NO. 3249

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

37 CFR 1.48(f)
ACKNOWLEDGEMENT LETTER

OC000000126344655

Date Mailed: 06/16/2021

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 06/14/2021.

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**.

/ggasgedom/

PART B–FEE(S) TRANSMITTAL

Complete and send this form, together with the 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, VA 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)

Brown Rudnick LLP
One Financial Center
Boston, Massachusetts 02111

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.

Sandra Richards	(Typed or printed name)
/Sandra Richards/	(Signature)
July 20, 2021	(Date)

APPLICATION NO	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO	CONFIRMATION NO
16/743,604	01/15/2020	Douglas A. Amorese	NUGN-001/03US 34079/516	3249

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

APPLN: TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1,200.00			$1,200.00	07/21/2021

EXAMINER	ART UNIT	CLASS-SUBCLASS
T. E. Strzelecka	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. Brown Rudnick LLP
2. Thomas C. Meyers
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)
TECAN GENOMICS, INC.	REDWOOD CITY, CA

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. 50-0369

5. Change of Entity Status (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29. 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.

☐ Applicant asserting small entity status. See 37 CFR 1.27. NOTE: If the application was previously under micro entity status, checking this box will be taken as a notification of loss of entitlement to micro entity status.

☐ Applicant changing to regular undiscounted fee status. NOTE: Checking this box will be taken as 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	/Thomas C. Meyers/	Date	July 20, 2021
Typed or printed name	Thomas C. Meyers	Registration No.	36,989

IDT Exhibit 1019.0506

Electronic Acknowledgement Receipt

EFS ID:	43293880
Application Number:	16743604
International Application Number:	
Confirmation Number:	3249
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/03US 34079/516
Receipt Date:	20-JUL-2021
Filing Date:	15-JAN-2020
Time Stamp:	14:31:06
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1200
RAM confirmation Number	E20217JE31210907
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	NUGN-001-03US-Issue-Fee-Transmittal.pdf	38322 76f082240f2ee380a937b2d427df39ce81bfff71	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30390 288ad9603f31e7a5f4d961d6eef31d5f361e8752	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	68712		

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.

Electronic Patent Application Fee Transmittal

Application Number:	16743604
Filing Date:	15-Jan-2020
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Filer:	Thomas C. Meyers/Sandra Richards
Attorney Docket Number:	NUGN-001/03US 34079/516

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

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

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

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.
16/743,604	08/24/2021	11098357	NUGN-001/03US 34079/516	3249

21710 7590 08/04/2021

BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

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

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

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

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

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Data Management (ODM) at (571)-272-4200.

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

Douglas A. Amorese, Los Altos, CA;
NUGEN TECHNOLOGIES, INC., Redwood City, CA;
Jonathan Scolnick, San Francisco, CA;
Ben Schroeder, San Mateo, CA;

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.