

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE**

ARCHERDX, LLC and THE GENERAL
HOSPITAL CORPORATION d/b/a
MASSACHUSETTS GENERAL HOSPITAL,

Plaintiffs,

V.

QIAGEN SCIENCES, LLC, QIAGEN LLC
f/k/a QIAGEN, INC., QIAGEN BEVERLY,
LLC f/k/a QIAGEN BEVERLY, INC.,
QIAGEN GAITHERSBURG, LLC f/k/a
QIAGEN GAITHERSBURG, INC., QIAGEN
GMBH, QIAGEN N.V., and
JONATHAN ARNOLD,

Defendants.

C.A. No. 18-1019-MN

PUBLIC VERSION FILED: March 8, 2021

**DECLARATION OF ERIC J. RUTT IN SUPPORT OF
PLAINTIFFS' MOTION FOR PARTIAL SUMMARY JUDGMENT
OF NO INVALIDITY AS TO U.S. PATENT NO. 10,450,597
AND TO STRIKE AND PRECLUDE TESTIMONY**

I, Eric J. Rutt, declare:

1. I am an attorney with the law firm Wolf Greenfield & Sacks, P.C. I have been admitted *pro hac vice* in this matter. I make this declaration in support of Plaintiffs' Motion for Partial Summary Judgment of No Invalidity as to U.S. Patent No. 10,450,597 and to Strike and Preclude Testimony. I am familiar with the facts set forth in this Declaration and could testify competently to the truth thereof if called as a witness.

2. Attached as Exhibit 1 is a true and correct copy of U.S. Patent No. 10,450,597 (“the ’597 Patent”).

3. Attached as Exhibit 2 is a true and correct copy of excerpts from the transcript of the February 9, 2021 deposition (Volume 1) of Dr. Michael Metzker.

4. Attached as Exhibit 3 is a true and correct copy of excerpts from the Expert Report of Dr. Michael Metzker Regarding Invalidity of U.S. Patent Nos. 10,017,810 and 10,450,597, dated October 7, 2020 (“Metzker Opening Report”).

5. Attached as Exhibit 4 is a true and correct copy of excerpts from the Rebuttal Expert Report of Dr. Niall Lennon, dated November 20, 2020 (“Lennon Rebuttal Report”).

6. Attached as Exhibit 5 is a true and correct copy of a publication authored by Newton et al, entitled “Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS),” which was attached as Exhibit LLL to the Reply Expert Report of Dr. Michael Metzker, and which was marked as Metzker Deposition Exhibit No. 10.

7. Attached as Exhibit 6 is a true and correct copy of excerpts from the Reply Expert Report of Dr. Michael Metzker Regarding Invalidity of U.S. Patent Nos. 10,017,810 and 10,450,597, dated December 23, 2020 (“Metzker Reply Report”).

8. Attached as Exhibit 7 is a true and correct copy of excerpts from the transcript of the February 10, 2021 deposition (Volume 2) of Dr. Michael Metzker.

9. Attached as Exhibit 8 is a true and correct copy of excerpts from the Expert Report of James A. McSwiggen, PhD, dated November 20, 2020 (“McSwiggen Report”).

10. Attached as Exhibit 9 is a true and correct copy of excerpts from the transcript of the January 22, 2021 deposition of James McSwiggen, PhD.

11. Attached as Exhibit 10 is a true and correct copy of excerpts of U.S. Patent No. 10,017,810 (“the ’810 Patent”).

12. Attached as Exhibit 11 is a true and correct copy of excerpts of U.S. Patent No. 9,487,828 (“the ’828 Patent”).

13. Attached as Exhibit 12 is a true and correct copy of excerpts of U.S. Patent App. Pub. No. 2013/0303461 (“the ’461 Publication”).

14. Attached as Exhibit 13 is a true and correct copy of excerpts of Manual of Patent Examining Procedure, Ninth Edition, June 2020 Publication of Revision 10.2019, which was accessed at <https://www.uspto.gov/web/offices/pac/mpep/index.html> on February 28, 2021.

15. Attached as Exhibit 14 is a true and correct copy of U.S. Provisional Patent Application No. 61/931,959 (“the ’959 Provisional”), which was produced in this litigation bearing production numbers ADX – 000365412 to ADX – 000365490.

16. Attached as Exhibit 15 is a true and correct copy of U.S. Provisional Patent Application No. 61/931,943 (“the ’943 Provisional”), which was produced in this litigation bearing production numbers ADX – 000365342 to ADX – 000365411.

17. Attached as Exhibit 16 is a true and correct copy of excerpts of the file history of U.S. Patent Application No. 14/605,363, which issued as the ’597 Patent, and which was produced in this litigation bearing production numbers ADX – 000485453 to ADX – 000491012.

18. Attached as Exhibit 17 is a true and correct copy of excerpts of the Noninfringement and No Trade Secret Misappropriation Expert Report of Dr. Michael Metzker Responding to the Expert Report of Niall Lennon, dated November 20, 2020 (“Metzker Rebuttal Report”).

19. Attached as Exhibit 18 is a true and correct copy of excerpts of Defendants’ Second Amended Invalidity Contentions, dated 10/2/2020, including excerpts of claim charts attached as Exhibits A-F.

20. Attached as Exhibit 19 is a true and correct copy of a summary of excerpts of Exhibit 14 (the '959 Provisional) and Exhibit 15 (the '943 Provisional).

I declare under penalty of perjury that the foregoing is true and correct. Executed on March 1, 2021.

/s/ Eric J. Rutt

Eric J. Rutt

EXHIBIT 1



US010450597B2

(12) **United States Patent**
Iafrate et al.

(10) **Patent No.:** **US 10,450,597 B2**
(45) **Date of Patent:** **Oct. 22, 2019**

(54) **METHODS OF PREPARING NUCLEIC ACIDS FOR SEQUENCING**

(71) Applicants: **The General Hospital Corporation**,
Boston, MA (US); **ArcherDX, Inc.**,
Boulder, CO (US)

(72) Inventors: **Anthony John Iafrate**, Newton, MA (US); **Long Phi Le**, Boston, MA (US);
Zongli Zheng, Boston, MA (US);
Jason Myers, Golden, CO (US);
Joshua Stahl, Boulder, CO (US)

(73) Assignees: **THE GENERAL HOSPITAL CORPORATION**, Boston, MA (US);
ARCHERDX, INC., Boulder, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 274 days.

(21) Appl. No.: **14/605,363**

(22) Filed: **Jan. 26, 2015**

(65) **Prior Publication Data**
US 2015/0211050 A1 Jul. 30, 2015

Related U.S. Application Data

(60) Provisional application No. 61/931,959, filed on Jan. 27, 2014.

(51) **Int. Cl.**
C12Q 1/68 (2018.01)
C12Q 1/6806 (2018.01)
C12Q 1/6809 (2018.01)

(52) **U.S. Cl.**
CPC **C12Q 1/6806** (2013.01); **C12Q 1/6809** (2013.01)

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,868,104 A 9/1989 Kurn et al.
5,827,658 A 10/1998 Liang
6,087,101 A 6/2000 Gruelich et al.
6,172,214 B1 1/2001 Brenner
6,576,448 B2 6/2003 Weissman et al.
6,632,611 B2 10/2003 Su et al.
6,833,246 B2 12/2004 Balasubramanian
7,214,490 B2 5/2007 Su et al.
7,244,559 B2 7/2007 Rothberg et al.
7,273,730 B2 9/2007 Du Breuil
7,824,856 B2 11/2010 Monforte
9,487,828 B2* 11/2016 Iafrate C12C 1/6874
10,017,810 B2 7/2018 Iafrate et al.
2002/0086317 A1 7/2002 Nagayama et al.
2003/0104432 A1 6/2003 Xu et al.
2003/0143553 A1 7/2003 Sommer
2007/0172824 A1* 7/2007 Chun C12Q 1/686
435/6.11
2009/0203085 A1* 8/2009 Kurn C12C 1/686
435/91.2

2010/0286143 A1 11/2010 Dias-Santagata et al.
2012/0122701 A1 5/2012 Ryan et al.
2013/0005585 A1* 1/2013 Anderson C12N 15/10
506/2
2013/0231253 A1 9/2013 Amorese et al.
2013/0303461 A1 11/2013 Iafrate et al.
2015/0011396 A1 1/2015 Schroeder et al.
2015/0140553 A1 5/2015 Cushing et al.
2015/0252361 A1 9/2015 Hayden et al.
2018/0127806 A1 5/2018 Stahl et al.
2018/0127807 A1 5/2018 Stahl et al.
2018/0155767 A1 6/2018 Myers et al.

FOREIGN PATENT DOCUMENTS

AL 2005203617 A1 9/2005
WO 97/23646 A1 7/1997
WO 97/23647 A1 7/1997
WO 98/28443 A1 7/1998
WO 99/42618 A1 8/1999
WO 00/43543 A1 7/2000
WO 00/70095 A2 11/2000
WO 01/12859 A2 2/2001
WO 01/20035 A2 3/2001
WO 2001/083696 A2 11/2001
WO 02/00938 A2 1/2002
WO 02/29117 A2 4/2002
WO 02/48402 A2 6/2002
WO 02/072772 A2 9/2002
WO 03/078645 A2 9/2003
WO 03/083435 A2 10/2003
WO 2004/011665 A2 2/2004
WO 2004/092418 A2 10/2004
WO 2005/065321 A2 7/2005
WO 2007/030759 A2 3/2007

(Continued)

OTHER PUBLICATIONS

Callaway, AS. et al. A sobemovirus coat protein gene complements long-distance movement of a coat protein-null dianthovirus. *Virology*, vol. 330, p. 186-195, 2004.*
Ali et al. "Sequence Analysis of TaphoA Insertion Sites in *Vibrio cholera* Mutants Defective in Rugose Polysaccharide Production". *Infection and Immunity* 68(12):6857-6864 (2000).
Baetens et al. "Applying Massive Parallel Sequencing to Molecular Diagnosis of Marfan and Loeys-Dietz Syndromes", *Human Mutation* 32:1053-1062 (2011).
Bohmer, "Novel application for isothermal nucleic acid sequence-based amplification (NASBA)", *Journal of Virological Methods* 158(1-2):199-201 (2009).
Chang et al., "Identification of a biomarker panel using a multiplex proximity ligation assay improves accuracy of pancreatic cancer diagnosis", *Journal of Translational Medicine* 7:105 (2009).

(Continued)

Primary Examiner — Suryaprabha Chunduru
(74) *Attorney, Agent, or Firm* — Nixon Peabody LLP;
David S. Resnick; Nicole D. Kling

(57) **ABSTRACT**

Aspects of the technology disclosed herein relate to methods for preparing and analyzing nucleic acids. In some embodiments, methods for preparing nucleic acids for sequence analysis (e.g., using next-generation sequencing) are provided herein.

29 Claims, 5 Drawing Sheets

Specification includes a Sequence Listing.

US 10,450,597 B2

Page 2

(56)

References Cited

FOREIGN PATENT DOCUMENTS

WO	2007/057652	A1	5/2007
WO	2007/136717	A1	11/2007
WO	2008/005459	A2	1/2008
WO	2008/093098	A2	8/2008
WO	WO2001/083696	A2 *	11/2008
WO	2009/102878	A2	8/2009
WO	2009/102896	A2	8/2009
WO	2009/117698	A2	9/2009
WO	2009/133466	A2	11/2009
WO	2009/148617	A2	12/2009
WO	2010/077288	A2	7/2010
WO	2010/083046	A2	7/2010
WO	2011/019964	A1	2/2011
WO	2011/032053	A1	3/2011
WO	2011/053987	A1	5/2011
WO	2011/156529	A2	12/2011
WO	2012/003374	A2	1/2012
WO	2012/040387	A1	3/2012
WO	2012/044956	A1	4/2012
WO	2012/064739	A2	5/2012
WO	2012/103154	A1	8/2012
WO	2013/059746	A1	4/2013
WO	2013/074833	A1	5/2013
WO	2013/112923	A1	8/2013
WO	2013/169339	A1	11/2013
WO	2013/191775	A2	12/2013
WO	2014/144092	A1	9/2014
WO	2014/150931	A1	9/2014
WO	2015/073711	A1	5/2015

OTHER PUBLICATIONS

Chen et al., "Allele-specific copy number profiling by next-generation DNA sequencing", *Nucleic Acids Research* 43(4):e23 (2015).

Chenchik, et al., "Full-Length cDNA Cloning and Determination of mRNA 5' and 3' Ends by Amplification of Adaptor-Ligated cDNA", *BioTechniques* 21(3):526-534 (1996).

Compton, "Nucleic acid sequence-based amplification", *Nature* 350(6313): 91-92 (1991).

Cushing et al., "RVD: a command-line program for ultrasensitive rare single nucleotide variant detection using targeted next-generation DNA resequencing", *BMC Research Notes* 6:206 (2013).

Dafforn et al., "Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis", *BioTechniques* 37(5):854-857 (2004).

Dahl et al., "Multigene amplification and massively parallel sequencing for cancer mutation discovery", *PNAS* 104(22):9387-9392 (2007).

Zhang et al., "Detecting simultaneous change-points in multiple sequences", *Biometrika* 97(3):631-645 (2010).

Zhang et al., "Titration-free 454 sequencing using Y adapters", *Nature Protocols* 6(9):1367-1376 (2011).

Lamant et al., "A New Fusion Gene TPM3-ALK in Anaplastic Large Cell Lymphoma Created by a (1;2)(q25;p23) Translocation", *Blood* 93(9):3088-3095 (1999).

Yuanxin et al., "T-linker-specific ligation PCT (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends", *Nucleic Acids Research* 31(12):e68 (2003).

Flaherty et al., "Ultrasensitive detection of rare mutations using next-generation targeted resequencing", *Nucleic Acids Research* 40(1):e2 (2012).

Fredriksson et al., "Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector", *Nucleic Acids Research* 35(7):e47 (2007).

Fredriksson et al., "Multiplexed protein detection by proximity ligation for cancer biomarker validation", *Nature Methods* 4(4):327-329 (2007).

Grace et al., "Degradable dUMP Outer Primers in Merged Tandem (M/T)-Nested PCR: Low- and Single-Copy DNA Target Amplification", *Analytical Biochemistry* 263(1):85-92 (1998).

Green et al., "Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma", *Blood* 121(9):1604-1611 (2013).

Grimes et al., "MendeLIMS: a web-based laboratory information management system for clinical genome sequencing", *BMC Bioinformatics* 15:290 (2014).

Guled et al., "Array comparative genomic hybridization analysis of olfactory neuroblastoma", *Modern Pathology* 21(6):770-778 (2008).

Head et al., "Method for improved Illumina sequencing library preparation using NuGEN Ovation RNA-Seq System", *Biotechniques* 50(3):177-181 (2011).

Hoeijmakers et al., "Linear amplification for deep sequencing", *Nature Protocols* 6(7):1026-1036 (2011).

Hoon et al., "Aptamer selection by high-throughput sequencing and informatic analysis", *Biotechniques* 51(6):413-416 (2011).

Hopmans et al., "A programmable method for massively parallel targeted sequencing", *Nucleic Acids Research* 42(10):e88 (2014).

Ji et al., "Data quality in genomics and microarrays", *Nat Biotechnol.* 24(9):1112-1113 (2006).

Ji et al., "Identification of a novel deletion mutant strain in *Saccharomyces cerevisiae* that results in a microsatellite instability phenotype", *Biodiscovery* 1:4 (2012).

Ji, "Improving bioinformatic pipelines for exome variant calling", *Genome Medicine* 4(1):7 (2012).

Ji et al., "Molecular Inversion Probe Analysis of Gene Copy Alterations Reveals Distinct Categories of Colorectal Carcinoma", *Cancer Res.* 66(16):7910-7919 (2006).

Ji et al., "Molecular inversion probe assay for allelic quantitation", *Methods Mol Biol.* 556:67-87 (2009).

Kim et al., "Genetic-based biomarkers and next-generation sequencing: the future of personalized care in colorectal cancer", *Per Med.* 8(3):331-345 (2011).

Kinde et al., "Detection and quantification of rare mutations with massively parallel sequencing", *PNAS* 108 (23):9530-9535 (2011).

Kurn et al., "Novel Isothermal, Linear Nucleic Acid Amplification Systems for Highly Multiplexed Applications", *Clinical Chemistry* 51(10):1973-1981 (2005).

Lam et al., "Performance comparison of whole-genome sequencing platforms", *Nat Biotechnol.* 30(1):78-82 (2011).

Lee et al., "Systematic genomic identification of colorectal cancer genes delineating advanced from early clinical stage and metastasis", *BMC Medical Genomics* 6:54 (2013).

Lin et al., "Reproducibility Probability Score—incorporating measurement variability across laboratories for gene selection", *Nature Biotechnology* 24(12):1476-1477 (2006).

Lishanski et al., "Branch migration inhibition in PCR-amplified DNA: homogeneous mutation detection", *Nucleic Acids Research* 28(9):e42 (2000).

MAQC Consortium, "The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements", *Nat Biotechnol.* 24(9):1151-1161 (2006).

Miotke et al., "Correction to High Sensitivity Detection and Quantitation of DNA Copy Number and Single Nucleotide Variants with Single Color Droplet Digital PCR", *Anal. Chem.* 87:3114 (2015).

Miotke et al., "High Sensitivity Detection and Quantitation of DNA Copy Number and Single Nucleotide Variants with Single Color Droplet Digital PCR", *Anal. Chem.* 86(5):2618-2624 (2014).

Mugasa et al., "Nucleic Acid Sequence-Based Amplification with Oligochromatography for Detection of *Trypanosoma brucei* in Clinical Samples", *Journal of Clinical Microbiology* 47(3):630-635 (2009).

Muralidharan et al., "A cross-sample statistical model for SNP detection in short-read sequencing data", *Nucleic Acids Research* 40(1):e5 (2012).

Myllykangas et al., "Classification of human cancers based on DNA copy number amplification modeling", *BMC Medical Genomics* 1:15 (2008).

Myllykangas et al., "DNA copy number amplification profiling of human neoplasms", *Oncogene* 25(55):7324-7332 (2006).

Myllykangas et al., "Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing", *Nat Biotechnol.* 29(11):1024-1027 (2011).

US 10,450,597 B2

Page 3

(56)

References Cited

OTHER PUBLICATIONS

- Mylykangas et al., "Integrated gene copy number and expression microarray analysis of gastric cancer highlights potential target genes", *Int. J. Cancer* 123(4):817-825 (2008).
- Mylykangas et al., "Manifestation, mechanisms and mysteries of gene amplifications", *Cancer Letters* 232(1):79-89 (2006).
- Mylykangas et al., "Novel high-throughput sequencing strategies in genetic diagnostics", *Duodecim* 129(2):141-148 (2013). English Abstract [Review Finnish].
- Mylykangas et al., "Specificity, selection and significance of gene amplifications in cancer", *Seminars in Cancer Biology* 17(1):42-55 (2007).
- Mylykangas et al., "Targeted deep resequencing of the human cancer genome using next-generation technologies", *Biotechnol Genet Eng Rev* 27:135-158 (2010).
- Mylykangas et al., "Targeted sequencing library preparation by genomic DNA circularization", *BMC Biotechnology* 11:122 (2011).
- Nadauld et al., "Quantitative and Sensitive Detection of Cancer Genome Amplifications from Formalin Fixed Paraffin Embedded Tumors with Droplet Digital PCR", *Transl Med (Sunnyvale)* 2(2) (2012).
- Natsoulis et al., "A Flexible Approach for Highly Multiplexed Candidate Gene Targeted Resequencing", *PLoS One* 6(6):e21088 (2011).
- Natsoulis et al., "Identification of Insertion Deletion Mutations from Deep Targeted Resequencing", *J Data Mining Genomics Proteomics* 4(3) (2013).
- Newburger et al., "The Human OligoGenome Resource: a database of oligonucleotide capture probes for resequencing target regions across the human genome", *Nucleic Acids Research* 40(Database issue):D1137-D1143 (2012).
- Scheinin et al., "CanGEM: mining gene copy number changes in cancer", *Nucleic Acids Research* 36(Database issue):D830-835 (2008).
- Schiffman et al., "Molecular inversion probes reveal patterns of 9p21 deletion and copy number aberrations in childhood leukemia", *Cancer Genet Cytogenet* 193(1):9-18 (2009).
- Schiffman et al., "Oncogenic BRAF Mutation with CDKN2A Inactivation is Characteristic of a Subset of Pediatric Malignant Astrocytomas", *Cancer Res* 70(2):512-519 (2010).
- Shendure et al., "Next-generation DNA sequencing", *Nature Biotechnology* 26(10):1135-1145 (2008).
- Singh et al., "Microarray-based comparison of three amplification methods for nanogram amounts of total RNA", *Am J Physiol Cell Physiol* 288(5):C1179-C1189 (2005).
- Sinkkonen et al., "Serial Analysis of Gene Expression in the Chicken Ootocyst", *JARO* 12(6):697-710 (2011).
- Tong et al., "Genome-scale identification of conditionally essential genes in *E. coli* by DNA microarrays", *Biochemical and Biophysical Research Communications* 322(1):347-354 (2004).
- Turner et al., "Gene Expression Profiling of RNA Extracted from FFPE Tissues: NuGEN Technologies' Whole-Transcriptome Amplification System", *Methods Mol Biol* 724:269-280 (2011).
- Xiao et al., "Sequential Amplification of Flanking Sequences by Y-shaped Adaptor Dependent Extension Using Multiple Templates", *33(1):85-90* (2007).
- Gould et al., "Analysis of sequence variation among smeDEF multi drug efflux pump genes and flanking DNA from defined 16S rRNA subgroups of clinical *Stenotrophomonas maltophilia* isolates.", *Journal of Antimicrobial Chemotherapy* 54(2):348-353 (2004).
- Liu et al., "High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences.", *Biotechniques Rapid Dispatches* 43(5):649-656 (2007).
- Wang et al., "Fusion primer and nested integrated PCR (FPNI-PCR): a new high-efficiency strategy for rapid chromosome walking or flanking sequence cloning.", *BMC Biotechnology* 11(1):109 (2011).
- Callaway et al., "A Sobemovirus coat protein gene complements long-distance movement of a coat protein-null Dianthovirus.", *Virology* 330(1):186-195 (2004).
- Frohman et al., "Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer.", *PNAS* 85(23):8998-9002 (1988).
- Kim et al., "FISH-negative cryptic PML-RARA rearrangement detected by long-distance polymerase chain reaction and sequencing analyses: a case study and review of the literature.", *Cancer Genetics and Cytogenetics* 203(2):278-283 (2010).

* cited by examiner

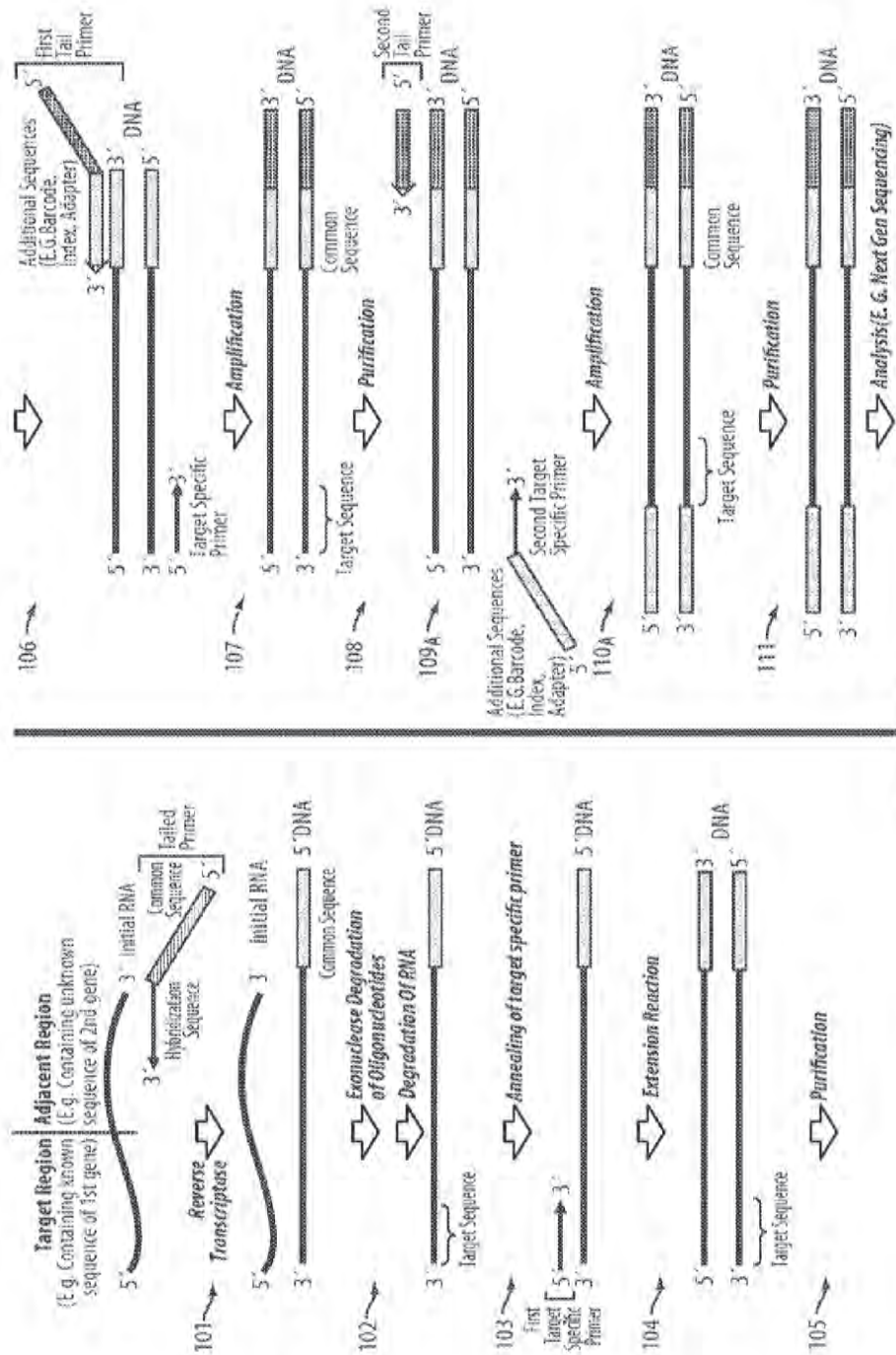


Fig. 1A

Fig. 1B

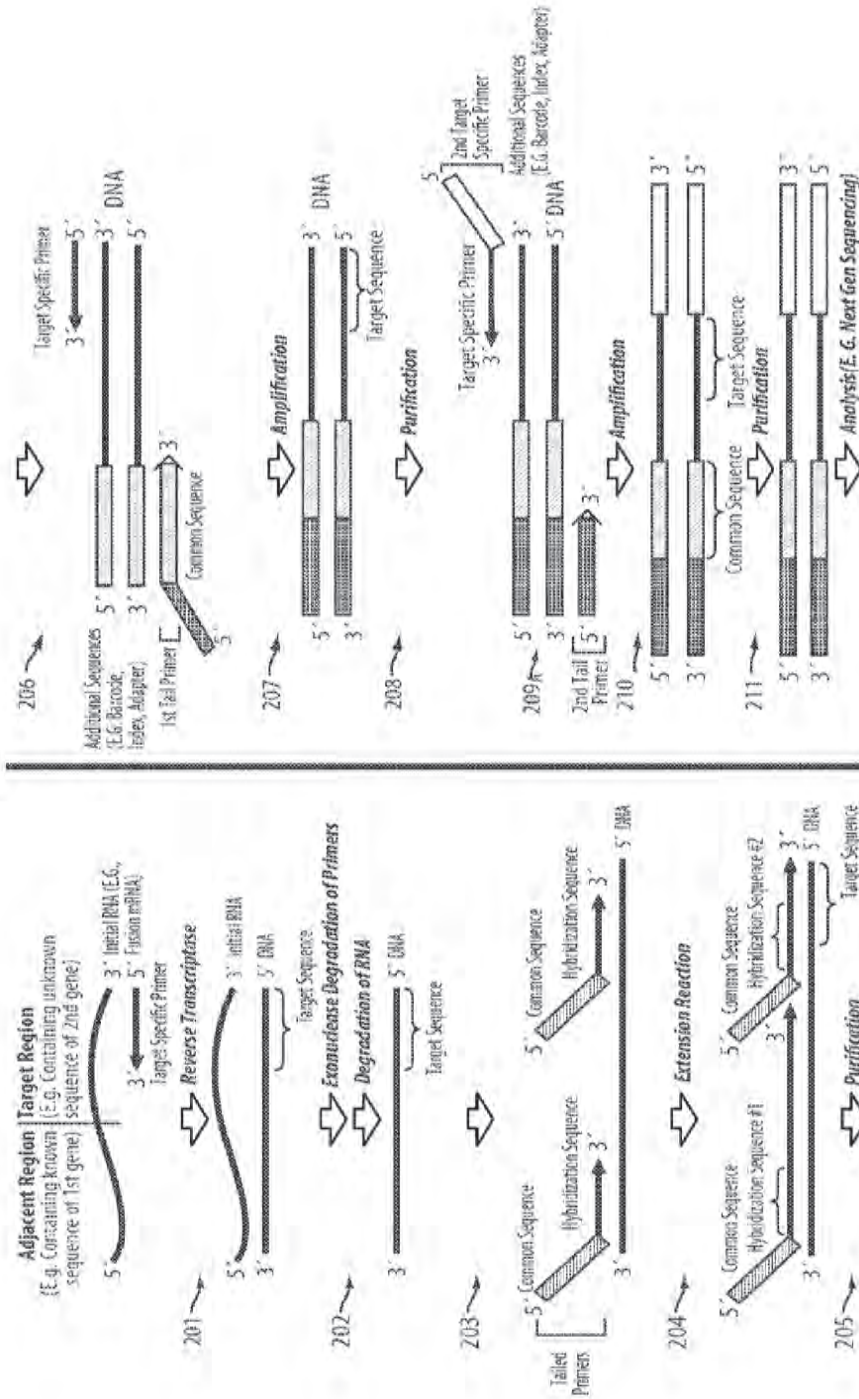


Fig. 2A

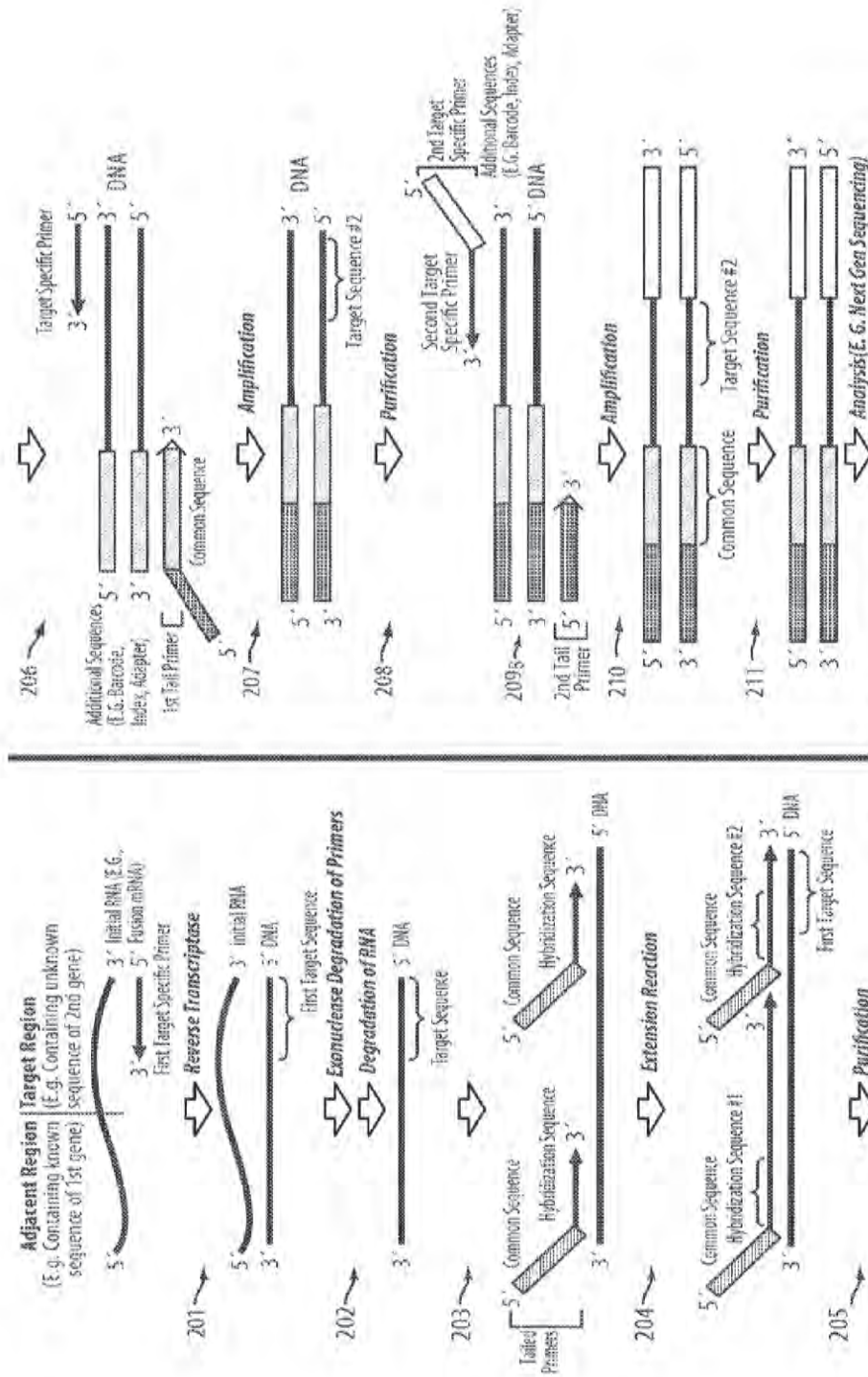


Fig. 2B

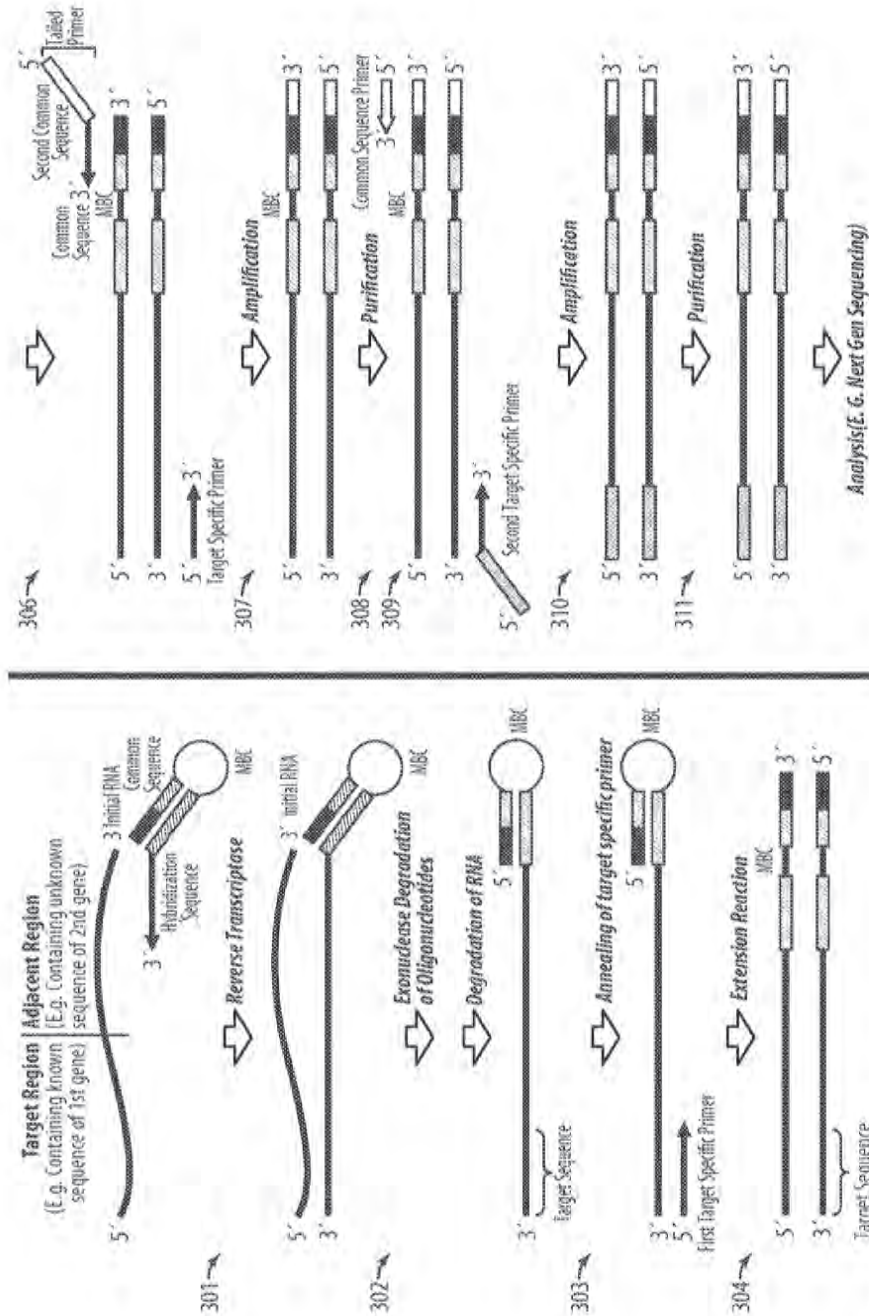


Fig. 3

US 10,450,597 B2

1

**METHODS OF PREPARING NUCLEIC ACIDS
FOR SEQUENCING****RELATED APPLICATIONS**

This application claims the benefit under 35 U.S.C. § 119 of U.S. provisional application 61/931,959 filed Jan. 27, 2014, the entire contents of which is incorporated herein by reference.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Mar. 13, 2015, is named 030258-079365-US SL.txt and is 31,392 bytes in size.

TECHNICAL FIELD

The technology described herein relates to methods of preparing and analyzing nucleic acids.

BACKGROUND

Target enrichment prior to next-generation sequencing is more cost-effective than whole genome, whole exome, and whole transcriptome sequencing and therefore more practical for broad implementation; both for research discovery and clinical applications. For example, high coverage depth afforded by target enrichment approaches enables a wider dynamic range for allele counting (in gene expression and copy number assessment) and detection of low frequency mutations, a critical feature for evaluating somatic mutations in cancer. Examples of current enrichment protocols for next generation sequencing include hybridization-based capture assays (TruSeq Capture, Illumina; SureSelect Hybrid Capture, Agilent) and polymerase chain reaction (PCR)-based assays (HaloPlex, Agilent; AmpliSeq, Ion Torrent; TruSeq Amplicon, Illumina; emulsion/digital PCR, Raindance). Hybridization-based approaches capture not only the targeted sequences covered by the capture probes but also near off-target bases that consume sequencing capacity. In addition, these methods are relatively time-consuming, labor-intensive, and suffer from a relatively low level of specificity.

SUMMARY

Aspects of the technology disclosed herein relate to methods for preparing and analyzing nucleic acids. In some embodiments, methods for preparing nucleic acids for sequence analysis (e.g., using next-generating sequencing) are provided herein. In some embodiments, technology described herein is directed to methods of determining nucleotide sequences of nucleic acids. In some embodiments, the methods described herein relate to enriching target nucleic acids prior to sequencing.

Aspects of the technology disclosed herein relate to methods of determining the nucleotide sequence contiguous to a known target nucleotide sequence. In some embodiments, the methods involve (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with an initial target-specific primer under hybridization conditions; (b) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid

2

molecule as a template; (c) contacting the product of step (b) with a population of tailed random primers under hybridization conditions; (d) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template; (e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer; (f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer; and (g) sequencing the amplified portion from step (f) using a first and second sequencing primer. In some embodiments, the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides. In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (e), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence identical to the tailed random primer. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

In some embodiments, the methods involve (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions; (b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template; (c) contacting the product of step (b) with an initial target-specific primer under hybridization conditions; (d) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid molecule as a template; (e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer; (f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer; and (g) sequencing the amplified portion from step (f) using a first and second sequencing primer. In some embodiments, the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides. In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical

US 10,450,597 B2

3

to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence identical to the tailed random primer. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

In some embodiments, the methods further involve a step of contacting the sample and/or products with RNase after extension of the initial target-specific primer. In some embodiments, the tailed random primer can form a hair-pin loop structure. In some embodiments, the initial target-specific primer and the first target-specific primer are identical. In some embodiments, the tailed random primer further comprises a barcode portion comprising 6-12 random nucleotides between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising 6-12 random nucleotides.

In some embodiments, the methods involve (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions; (b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template; (c) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer; (d) amplifying a portion of the amplicon resulting from step (c) with a second tail primer and a second target-specific primer; and (e) sequencing the amplified portion from step (d) using a first and second sequencing primer. In some embodiments, the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer; a middle barcode portion comprising; and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides. In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence identical to the tailed random primer. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer. In some embodiments, the each tailed random primer further comprises a spacer nucleic acid sequence between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising about 6 to about 12 random nucleotides. In certain embodiments, the unhybridized primers are removed from the reaction after an extension step. In some embodiments, the second tail primer is nested with respect to the first tail primer by at least 3 nucleotides. In certain embodiments, the first target-specific primer further comprises a 5' tag sequence portion comprising a nucleic acid sequence of high GC content which is not substantially complementary to or substantially identical to

4

any other portion of any of the primers. In some embodiments, the second tail primer is identical to the full-length first sequencing primer. In certain embodiments, the portions of the target-specific primers that specifically anneal to the known target will anneal specifically at a temperature of about 65° C. in a PCR buffer. In some embodiments, the sample comprises genomic DNA. In some embodiments, the sample comprises RNA and the method further comprises a first step of subjecting the sample to a reverse transcriptase regimen. In certain embodiments, the nucleic acids present in the sample have not been subjected to shearing or digestion. In some embodiments, the sample comprises single-stranded gDNA or cDNA. In certain embodiments, the reverse transcriptase regimen comprises the use of random hexamers. In some embodiments, a gene rearrangement comprises the known target sequence. In certain embodiments, the gene rearrangement is present in a nucleic acid selected from the group consisting of: genomic DNA; RNA; and cDNA. In some embodiments, the gene rearrangement comprises an oncogene. In certain embodiments, the gene rearrangement comprises a fusion oncogene. In some embodiments, the nucleic acid product is sequenced by a next-generation sequencing method. In certain embodiments, the next-generation sequencing method comprises a method selected from the group consisting of: Ion Torrent, Illumina, SOLiD, 454; Massively Parallel Signature Sequencing solid-phase, reversible dye-terminator sequencing; and DNA nanoball sequencing. In certain embodiments, the first and second sequencing primers are compatible with the selected next-generation sequencing method. In some embodiments, the method comprises contacting the sample, or separate portions of the sample, with a plurality of sets of first and second target-specific primers. In certain embodiments, the method comprises contacting a single reaction mixture comprising the sample with a plurality of sets of first and second target-specific primers. In some embodiments, the plurality of sets of first and second target-specific primers specifically anneal to known target nucleotide sequences comprised by separate genes. In certain embodiments, at least two sets of first and second target-specific primers specifically anneal to different portions of a known target nucleotide sequence. In some embodiments, at least two sets of first and second target-specific primers specifically anneal to different portions of a single gene comprising a known target nucleotide sequence. In certain embodiments, at least two sets of first and second target-specific primers specifically anneal to different exons of a gene comprising a known nucleotide target sequence. In some embodiments, the plurality of first target-specific primers comprise identical 5' tag sequence portions. In certain embodiments, each tailed random primer in a population of tailed random primers further comprises an identical sample barcoding portion. In some embodiments, multiple samples are each contacted with a separate population of tailed random primers with a sample barcoding portion; wherein each population of tailed random primers has a distinct sample barcoding portion; and wherein the samples are pooled after step (b). In certain embodiments, each amplification step comprises a set of cycles of a PCR amplification regimen from 5 cycles to 20 cycles in length. In some embodiments, the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of from about 61 to 72° C. In some embodiments, the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of about 65° C. In certain embodiments, the target nucleic acid

US 10,450,597 B2

5

molecule is from a sample, optionally which is a biological sample obtained from a subject. In some embodiments, the sample is obtained from a subject in need of treatment for a disease associated with a genetic alteration. In certain embodiments, the disease is cancer. In some embodiments, the sample comprises a population of tumor cells. In certain embodiments, the sample is a tumor biopsy. In some embodiments, the cancer is lung cancer. In certain embodiments, a disease-associated gene comprises the known target sequence. In some embodiments, a gene rearrangement product in the sample comprises the known target sequence. In certain embodiments, the gene rearrangement product is an oncogene.

Aspects of the technology disclosed herein relate to methods of preparing nucleic acids for analysis. In some embodiments, the methods involve method (a) contacting a nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer; and (b) contacting a nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers, in which an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers. In some embodiments, the target nucleic acid is a ribonucleic acid. In certain embodiments, the target nucleic acid is a deoxyribonucleic acid. In some embodiments, steps (a) and (b) are performed sequentially. In certain embodiments, the nucleic acid template in step (a) comprises an extension product resulting from the hybridization and extension of the at least one of the plurality of different primers in step (b). In some embodiments, the nucleic acid template in step (b) comprises an extension product resulting from the hybridization and extension of the target-specific primer in step (a). In certain embodiments, the target nucleic acid is a messenger RNA encoded from a chromosomal segment that comprises a genetic rearrangement. In some embodiments, the target nucleic acid is a chromosomal segment that comprises a portion of a genetic rearrangement. In certain embodiments, the genetic rearrangement is an inversion, deletion, or translocation. In some embodiments, the methods further involve amplifying the extension product. In certain embodiments, the methods further involve contacting the extension product or amplified extension product with an immobilized oligonucleotide under conditions in which hybridization occurs between the extension product and immobilized oligonucleotide. In certain embodiments, the target nucleic acid comprises a target portion having a known sequence and a flanking portion having an unknown sequence. In some embodiments, different hybridization sequences are complementary to the flanking portion. In certain embodiments, the target-specific hybridization sequence is complementary to the target portion. In some embodiments, the target-specific primer further comprises, 5' to the target-specific hybridization sequence, at least one of an index sequence, a barcode sequence and an adaptor sequence. In certain embodiments, the common sequence comprises at least one of an index sequence, barcode sequence and an adaptor sequence. In

6

some embodiments, the adaptor sequence is a cleavable adaptor sequence for immobilizing oligonucleotides in a flow cell.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A and 1B depict non-limiting embodiments of a work flow for amplifying and sequencing target nucleic acids that are flanked by a 3' unknown fusion partner, as described herein.

FIGS. 2A and 2B depict non-limiting embodiments of a work flow for amplifying sequencing target nucleic acids that are flanked by a 5' unknown fusion partner, as described herein.

FIG. 3 depicts a non-limiting embodiment of a work flow for amplifying sequencing target nucleic acids that are flanked by a 3' unknown fusion partner using an oligonucleotide comprising a hairpin, as described herein.

DETAILED DESCRIPTION

Aspects of the technology disclosed herein relate to methods for preparing and analyzing nucleic acids. In some embodiments, methods provided herein are useful for determining unknown nucleotide sequences contiguous to (adjacent to) a known target nucleotide sequence. Traditional sequencing methods generate sequence information randomly (e.g., "shotgun" sequencing) or between two known sequences which are used to design primers. In contrast, methods described herein, in some embodiments, allow for determining the nucleotide sequence (e.g., sequencing) upstream or downstream of a single region of known sequence with a high level of specificity and sensitivity. Accordingly, in some embodiments, methods provided herein are useful for determining the sequence of fusions (e.g., fusion mRNAs) that result from gene arrangements (e.g., rearrangements that give rise to cancer or other disorders). In some embodiments, methods provided herein for preparing nucleic acids for analysis (e.g., for sequencing) involve a first round of extension using a target-specific primer that targets a known sequence of a target nucleic acid (e.g., a known sequence of a 1st gene) followed by a second round of extension that involves the use of a heterogeneous population of tailed primers (e.g., tailed random primers) that include tailed primers that have hybridization sequences that are complementary with an unknown sequence adjacent to the known sequence in the target nucleic acid. In some embodiments, the tail region of tailed primers includes barcode or index sequences that facilitate multiplex amplification and enrichment of target nucleic acids.

In some aspects of the technology disclosed herein methods are provided of preparing nucleic acids for analysis that involve (a) contacting a nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer and (b) contacting a nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers, in which an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of

US 10,450,597 B2

7

different primers. In some embodiments, steps (a) and (b) above are performed sequentially. In some embodiments, the nucleic acid template in step (a) comprises an extension product resulting from the hybridization and extension of the at least one of the plurality of different primers in step (b). In some embodiments, the nucleic acid template in step (b) comprises an extension product resulting from the hybridization and extension of the target-specific primer in step (a).

In some embodiments, methods are provided for preparing nucleic acids that have a target region 5' to an adjacent region (e.g., an adjacent region of unknown sequence). In some embodiments, methods provided herein can be accomplished using one or more rounds of PCR.

For example, FIGS. 1A-B present schematics of exemplary methods of amplifying target nucleic acids that have a known target region 5' to an adjacent region (e.g., for purposes of sequencing the adjacent region). At step 101, initial RNA is obtained or provided in a sample and is used as a template. RNA template is exposed to a plurality of tailed primers (e.g., tailed random primers) that comprise a common sequence that is 5' to different hybridization sequences and shared between all of the tailed primers of the population. In some embodiments, at least one primer hybridizes to an RNA molecule and primes a reverse transcriptase reaction to produce a complementary DNA strand. In step 102, unhybridized oligonucleotides are degraded (e.g., enzymatically, e.g., by an exonuclease). In step 102, RNA template is degraded from the complementary DNA strand.

In some embodiments, a tailed primer is provided that hybridizes to the poly-A tail of an RNA molecule. In some embodiments, the sequence of a primer is provided that hybridizes to the poly-A tail comprises a poly-dT (e.g., a 3' positioned stretch of 2 dTs, 3 dTs, 4 dTs, 5 dTs, 6 dTs, 7 dTs, 8 dTs, 9 dTs, 10 dTs, or more.). In some embodiments, a plurality of tailed primers are provided each of which comprises a common sequence. In some embodiments, a plurality of tailed primers is provided each of which further comprises a barcode or index sequence.

It should be appreciated that in methods disclosed herein an RNA template may be degraded by any appropriate method, including, for example, by enzymatic degradation (e.g., using RNaseH, Uracyl glycosylase, etc.), by hydrolysis (e.g., by exposing the RNA to relatively high pH conditions (e.g., pH 10, pH 11, pH 12), etc. In some embodiments, hydrolyzing RNA by exposure to relatively high pH is advantageous because it is relatively inexpensive (compared with certain other methods, e.g., certain enzymatic methods) and because it destroys both RNA and DNA:RNA hybrids. In some embodiments, RNA is degraded by hydrolysis caused by exposure to relatively high pH conditions at a relatively high temperature (e.g., temperature greater than 60° C. (e.g., 60° C. to 95° C.)). In some embodiments, the use of relatively high temperatures is advantageous because it heat inactivates enzymes used in prior preparative steps (e.g., RT enzymes). In some embodiments, an initial nucleic acid may be DNA that is obtained or provided in a sample and is used as a template. In such embodiments, steps 101 and 102 may be omitted.

In step 103, DNA molecules produced by reverse transcription are contacted by one or more initial target-specific primers which may or may not be the same as the first target-specific primer. In step 104, hybridization of the initial target-specific primer to a portion of the target nucleic acid (the "target sequence") primes an extension reaction using a DNA molecule as a template to produce a complementary DNA strand. Extension products are purified in step

8

105. However, in some embodiments, DNA produced in step 104 may be amplified directly, e.g., by PCR, without purification.

In step 106, DNA molecules are contacted by a first target-specific primer and a first tail primer. The first target-specific primer hybridizes to a portion of the target nucleic acid. In some embodiments, pools of different first target-specific primers can be used that hybridize to different portions of a target nucleic acid. In some embodiments, use of different target specific primers can be advantageous because it allows for generation of different extension products having overlapping but staggered sequences relative to a target nucleic acid. In some embodiments, different extension products can be sequenced to produce overlapping sequence reads. In some embodiments, overlapping sequence reads can be evaluated to assess accuracy of sequence information, fidelity of nucleic acid amplification, and/or to increase confidence in detecting mutations, such as detecting locations of chromosomal rearrangements (e.g., fusion breakpoints). In some embodiments, pools of different first target-specific primers can be used that hybridize to different portions of different target nucleic acids present in sample. In some embodiments, use of pools of different target-specific primers is advantageous because it facilitates processing (e.g., amplification) and analysis of different target nucleic acids in parallel. In some embodiments, up to 2, up to 3, up to 4, up to 5, up to 6, up to 7, up to 8, up to 9, up to 10, up to 15, up to 20, up to 100 or more pools of different first target-specific primers are used. In some embodiments, 2 to 5, 2 to 10, 5 to 10, 5 to 15, 10 to 15, 10 to 20, 10 to 100, 50 to 100, or more pools of different first target-specific primers are used.

In FIGS. 1A and 1B, a first tail primer hybridizes to at least a portion of a DNA molecule provided by the tail portion of the tailed primer of step 101. In some embodiments, the first tail primer hybridizes to the common sequence provided by the tail of the one or more primers of step 101. In some embodiments, a nested target specific primer (nested with respect to the target specific primer of step 103) is used in step 106. In some embodiments, a first tail primer may comprise an additional sequence 5' to the hybridization sequence that may include barcode, index, adapter sequences, or sequencing primer sites, for example. In step 107, hybridization of the first target-specific primer and the first tail nucleic acid molecule allows for amplification of a product in a polymerase chain reaction (PCR). In some embodiments, amplified products are purified in step 108.

In some embodiments, the ssDNA product of step 103 is amplified directly (e.g., by PCR) rather than performing the extension reaction of step 104 and purification of step 105. Similarly, in some embodiments of any of the methods disclosed herein, ssDNA products may be amplified directly (e.g., by PCR) rather than performing an extension reaction to produce dsDNA prior to purification and/or PCR. In some embodiments, first and second tail primers can be incorporated during a PCR. In some embodiments, primers (e.g., first tail primers, target specific primers) are used in a PCR or extension reaction and then excess primer is removed using a single stranded nuclease. Subsequent rounds of PCR or extension may be performed using different primers (e.g., second tail primers or nested primers or a second target specific primer) to incorporate different sequences into the resulting products.

In some embodiments, an exonuclease (e.g., Exol) may be used to degrade single-stranded DNA. In some embodiments, an exonuclease is used to degrade ssDNA and the

US 10,450,597 B2

9

amplified product is processed directly according to steps 109_{A-B} and 110_{A-B}, without purification at step 108.

In FIG. 1A, at step 109_A, amplified DNA products of step 107 (e.g., as purified in step 108) are contacted with a second target-specific primer and a second tail primer. In some embodiments, the second target-specific primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first target-specific primer such that the reactions are nested. In some embodiments, nesting of the second target-specific primer relative to the first target-specific primer may improve specificity of the hybridization reaction. In some embodiments, the second target-specific primer may comprise an additional sequence 5' to the hybridization sequence that may include barcode, index, adapter sequences, or sequencing primer sites, for example. In step 110_A, the amplified DNA products of step 107 (e.g., as purified in step 108) are amplified by PCR in which the extensions are primed by the second target-specific primer and a second tail primer. In some embodiments, a portion of the amplified product from step 107 is further amplified. In some embodiments, a third primer is used that hybridizes to the common tail in the second target specific primer and adds additional sequences such as a barcodes, adapters, etc.

In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such that the reactions are nested. In such embodiments, a portion of the product from step 106 is amplified. In some embodiments, the second tail primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. Hybridization of the second target-specific primer and the second tail primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction.

In some embodiment, the first target-specific primer of step 106 may be used with a second target-specific primer. In such embodiments, hybridization of the first target-specific primer and the second target-specific primer allows for amplification of product in a polymerase chain reaction (PCR). In some embodiments, steps 108-110 may be omitted. The amplification products are purified in reaction 111 and ready for analysis. For example, products purified in step 111 can be sequenced (e.g., using a next generation sequencing platform.) In some embodiments the first target-specific primer or second target-specific primer may comprise additional sequences 5' to a hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites.

In some embodiments, as depicted in FIG. 1B, in step 109_B, DNA products of step 107 (e.g., as purified in step 108) are contacted with a second target-specific primer and a second tail primer. The second target-specific primer is further contacted by an additional primer (e.g., a primer having 3' sequencing adapter/index sequences) that hybridizes with the common sequence of the second target-specific primer. In some embodiments the additional primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. In some embodiments, the additional primer is a generic sequencing adapter/index primer. In some embodiments, the second target-specific primer may be nested relative to the target-specific

10

primer used in step 107. In step 110_B, the DNA products of step 107 (e.g., as purified in step 108) are amplified by PCR in which the extensions are primed by the second target-specific primer and a second tail primer. Hybridization of the second target-specific primer, the additional primer, and the second tail primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction. In such embodiments, a portion of the amplified product from step 108 is amplified.

In some embodiment, the first tail primer of step 106 may be used with a second target-specific primer. In such embodiments, hybridization of the first tail primer and the second target-specific primer allows for amplification of product in a polymerase chain reaction (PCR), optionally with an additional primer (e.g., a primer having 3' sequencing adapter/index sequences) that hybridizes with the common sequence of the second target-specific primer. In some embodiments, steps 108-110 may be omitted.

The products are purified in reaction 111 and ready for analysis. For example, products purified in step 111 can be sequenced (e.g., using a next generation sequencing platform).

In some embodiments, steps 101-104, 106-107, and 109-110 are performed consecutively in a single reaction tube without any intervening purification steps. In some embodiments, all of the components involved in steps 101-104, 106-107, 109-110 are present at the outset and throughout the reaction. In some embodiments, steps 101-104 are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps 101-104 are present at the outset and throughout the reaction. In some embodiments, steps 106-107 are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps 106-107 are present at the outset and throughout the reaction. In some embodiments, steps 109_A-110_A or 109_B-110_B are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps 109_A-110_A or 109_B-110_B are present at the outset and throughout the reaction.

In some embodiments, methods are provided for preparing nucleic acids that have a target region 3' to an adjacent region (e.g., an adjacent region of unknown sequence content). For example, FIG. 2 presents a schematic of an exemplary method of amplifying and sequencing target nucleic acids that have a known target region 3' to an adjacent region. In FIGS. 2A-B, an initial RNA (e.g., a fusion mRNA) is obtained or provided in a sample and is used as a template for the proceeding method. At step 201, the RNA template is exposed to one or more initial target-specific primers that hybridize to one or more target nucleotide sequences and function to prime a reverse transcription reaction such that a complementary DNA molecule is produced using the initial RNA as a template. In some embodiments, an initial target-specific primer hybridizes to the poly-A tail of an RNA template. In some embodiments, the sequence of the primer that hybridizes to the poly-A tail comprises a poly-dT (e.g., a 3' positioned stretch of 2 dTs, 3 dTs, 4 dTs, 5 dTs, 6 dTs, 7 dTs, 8 dTs, 9 dTs, 10 dTs, or more). In step 202 the unhybridized primers are degraded (e.g., enzymatically, e.g., by an exonuclease). In step 202, RNA template is degraded from the complementary DNA strand (e.g., enzymatically, e.g., by RNaseH).

In step 203, DNA molecules that were generated by reverse transcription are contacted by a heterogeneous population of tailed primers (e.g., tailed random primers). In some embodiments, the tail portion of each of the tailed primers is a shared or common sequence, identical between

US 10,450,597 B2

11

all primers of the population of tailed primers. In some embodiments, at least one primer comprises a hybridization sequence that is complementary to and hybridizes to the target acid template. In step 204, tailed primers that are hybridized with the template nucleic acid are extended in a template-dependent extension reaction to produce complementary DNA strands that incorporate the tailed primer sequence and the template sequence. The resulting double stranded DNA product is purified in step 205.

In step 206, DNA products purified in step 205 are contacted with a first target-specific primer and a first tail primer. The first target-specific primer hybridizes to a target sequence (region) of the DNA. The first tail primer hybridizes to a portion of the DNA molecule characteristic of the tail of the tailed primers of step 203. In some embodiments, the first tail primer hybridizes to the common sequence provided by the tail of the one or more primers of step 203. In some embodiments, a first tail primer comprises a barcode or index sequence. In step 207 hybridization of the first target-specific primer and the first tail primer facilitates exponential amplification of a portion of the target nucleic acid molecule in an amplification reaction (e.g., a PCR reaction). In some embodiments, the first tail primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. Amplified products are purified in step 208. In some embodiments, the purification in step 208 is skipped. For example, in some embodiments, primers (e.g., second target specific primers and 2nd tail primers) are used in a PCR or extension reaction and then excess primer is removed using a single stranded DNA nuclease. Subsequent rounds of PCR or extension may be performed using different primers to incorporate different sequences into the resulting products.

In some embodiments, as depicted in FIG. 2A at step 209_A, DNA molecules produced in step 207 (e.g., as purified in step 208) are contacted with a second target-specific primer and a second tail primer. In some embodiments, the second target-specific primer is nested relative to the target specific primer used in step 207. In some embodiments, at least a portion of the product from step 209_A is amplified.

In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such that the reactions are nested. In some embodiments, at least a portion of the product from step 209_A is amplified.

In some embodiments, first tail primer of step 206 may be used with a second target-specific primer. In such embodiments, hybridization of the first tail primer and the second target-specific primer allows for amplification of product in a polymerase chain reaction (PCR). In such embodiments, steps 208-210 may be omitted.

In some embodiments, as depicted in FIG. 2B at step 209_B, DNA molecules produced in step 207 (e.g., as purified in step 208) are contacted with a second target-specific primer and a second tail primer wherein the second target-specific primer hybridizes to a target sequence that is present within the template DNA molecule 3' of the sequence of the first target-specific primer such that the reactions are nested. In some embodiments, at least a portion of the product from step 209_B is amplified.

In some embodiments, the second target-specific primer is further contacted by an additional primer (e.g., a primer having 3' sequencing adapter/index sequences) that hybrid-

12

izes with the common sequence of the second target-specific primer. In some embodiments the additional primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. In some embodiments, the additional primer is a generic sequencing adapter/index primer. In such embodiments, hybridization of the second target-specific primer, the additional primer, and a second tail primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction.

In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such that the reactions are nested. In some embodiments, at least a portion of the product from step 209_B is amplified.

In some embodiments, the second tail primer may comprise additional sequence 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. In step 210, hybridization of the second target-specific primer and the second tail primer facilitates exponential amplification of a portion of the target nucleic acid molecule in an amplification reaction (e.g., a PCR reaction). The amplification product of step 210 is purified in reaction 211 and ready for analysis. For example, products purified in step 211 can be sequenced (e.g., using a next generation sequencing platform.)

In some embodiments, steps 201-204 are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps 201-204 are present at the outset and throughout the reaction. In some embodiments, steps 206-207 are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps 206-207 are present at the outset and throughout the reaction. In some embodiments, steps 209_A-210 or 209_B-210 are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps 209_A-210 or 209_B-210 are present at the outset and throughout the reaction.

In some embodiments, methods provided herein involve use of random sequences as molecular barcodes. In some embodiments, molecular barcodes are built into primers (e.g., RT primers, target-specific primers, extension sequence primers) such that each individual molecule produced by a primer obtains a unique barcode tag. Thus, in some embodiments, the molecular barcode tag permits a determination of whether a sequenced molecular is unique.

In some embodiments, molecular barcodes may be used to silence sequencing errors, improve confidence calling of fusions or other mutations, and improved detection limits.

In some embodiments, methods are provided for preparing nucleic acids that have a target region 5' to an adjacent region (e.g., an adjacent region of unknown sequence) using an oligonucleotide comprising a hairpin. In some embodiments, the oligonucleotide may have a structure that is not a hairpin, for example, the oligonucleotide may be linear. For example, FIG. 3 presents a schematic of an exemplary method for amplifying target nucleic acids that have a known target region 5' to an adjacent region (for purposes of sequencing the adjacent region). At step 301, an initial RNA is obtained or provided in a sample and is used as a template for the proceeding method. The RNA template is exposed to a plurality of hairpin primers (e.g., random primers with a hairpin tail) that comprise a hairpin sequence that is 5' to different hybridization sequences and shared between all of

US 10,450,597 B2

13

the primers of the population. The hairpin sequence comprises two complementary common sequences that flank a molecular barcode sequence (MBC). The complementary common sequences base pair to form the stem-loop hairpin structure and protect the MBC sequence. In some embodiments, the plurality of primers is in another structure (not a hairpin) and comprises two complementary common sequences that flank a molecular barcode sequence (MBC). In some embodiments, at least one primer hybridizes to the RNA molecule and primes a reverse transcriptase reaction to produce a complementary DNA strand. In some embodiments, a primer hybridizes to the poly-A tail of the RNA molecule. In some embodiments, the sequence of the primer that hybridizes to the poly-A tail comprises a poly-dT (e.g., a 3' positioned stretch of 2 dTs, 3 dTs, 4 dTs, 5 dTs, 6 dTs, 7 dTs, 8 dTs, 9 dTs, 10 dTs, or more.). In step 302, any unhybridized oligonucleotides are enzymatically degraded (e.g., by an exonuclease). Also in step 302, the RNA template is enzymatically degraded from the complementary DNA strand (e.g., by RNaseH).

In step 303, DNA molecules produced by reverse transcription are contacted by one or more initial target-specific primers which may or may not be the same as the first target-specific primer. In step 304, hybridization of the first target-specific primer to a portion of the target nucleic acid primes an extension reaction using the DNA molecule as a template to produce a complementary DNA strand. In some embodiments, synthesis of the complementary DNA strand may reduce or eliminate hairpin formation of the complementary common sequences. Extension products are purified in step 305.

In step 306, DNA molecules are contacted by a first target-specific primer and a tailed primer. The first target-specific primer hybridizes to a portion of the target nucleic acid. The first tail primer hybridizes to a portion of the DNA molecule provided by the common sequence involve in hairpin formation step 301. In some embodiments, a nested target-specific primer (e.g., nested with respect to the target-specific primer of step 303) is used in step 306. In some embodiments, the first tailed primer may comprise an additional sequence 5' to the hybridization sequence that may include barcode, index, adapter sequences, or sequencing primer sites, for example. In step 307, hybridization of each of the first target-specific primer and the tailed primer allows for amplification of a portion of the target nucleic acid molecule in a polymerase chain reaction (PCR). In some embodiments, amplified products are purified in step 308.

In step 309, amplified DNA products (e.g., those purified in step 308) are contacted with a second target-specific primer and a common sequence primer. In some embodiments, the second target-specific primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first target-specific primer such that the reactions are nested. In some embodiments, the common sequence primer hybridizes with a sequence provided by the first tail primer in step 306. In some embodiments, in step 310, DNA products purified in step 308 are amplified by PCR in which the extensions are primed by a second target-specific primer and a common sequence primer. In some embodiments, amplified product from step 308 may be amplified.

In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such

14

that the reactions are nested. In such embodiments, a portion of the product from step 308 is amplified. In some embodiments, the common sequence primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. Hybridization of the second target-specific primer and the common sequence primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction. In some embodiments, products are purified in reaction 311 useful for analysis. For example, products purified in step 311 can be sequenced (e.g., using a next generation sequencing platform.)

In some embodiments, all of the components involved in steps 301-311 are present at the outset and throughout the reaction. In some embodiments, steps 301-304, 306-307, and 309-310 are performed consecutively in a single reaction tube without any intervening purification steps. In some embodiments, steps 301-304 are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps 301-304 are present at the outset and throughout the reaction. In some embodiments, steps 306-307 are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps 306-307 are present at the outset and throughout the reaction. In some embodiments, steps 309-310 are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps 309-310 are present at the outset and throughout the reaction.

In some embodiments, methods are provided herein that involve determining the nucleotide sequence contiguous to (adjacent to) a known target nucleotide sequence. In some embodiments, methods comprise contacting a target nucleic acid molecule comprising the known target nucleotide sequence with an initial target-specific primer under suitable hybridization conditions. In some embodiments, the methods further comprise maintaining the target nucleic acid molecule under conditions that promote extension of the hybridized initial target-specific primer (e.g., using the target nucleic acid molecule as a template), thereby producing a first extension product. In some embodiments, the methods further comprise contacting the extension product with a population of tailed random primers under suitable hybridization conditions. In some embodiments, the methods further comprise maintaining the extension product under conditions that promote extension of a hybridized tailed random primer using the portion of the target nucleic acid molecule downstream of the site of hybridization as a template, thereby producing a second extension product. In some embodiments, the methods further comprise amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer, thereby producing a first amplicon. In some embodiments, the methods further comprise amplifying a portion of the amplicon with a second tail primer and a second target-specific primer, thereby producing a second amplicon.

In some embodiments, one or more target-specific primers used in the methods may be nested with respect to one or more other target-specific primers. For example, in some embodiments, a second target-specific primer is internal to a first target-specific primer. In some embodiments, target-specific primers are the same. In some embodiments, target-specific primers are nested but overlapping with respect to target complementarity. In some embodiments, target-specific primers are nested and non-overlapping. In some embodiments, combinations of identical and nested target

US 10,450,597 B2

15

specific primers are used in the same or different amplification steps. In some embodiments, nesting of primers increases target specificity. In some embodiments, the methods further comprise sequencing the second amplicon using a first and second sequencing primer. In some embodiments, the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleotide sequence identical to a first sequencing primer and a 3' nucleotide comprising from random nucleotides (e.g., about 6 to about 12 random nucleotides). In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known nucleotide sequence of the target nucleic acid at an appropriate annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the first amplicon, and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence identical to the common sequence of the tail of the tailed random primer. In some embodiments, the common sequence on the tailed random primer is the exact match of the common sequence on the first tail primer. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

As used herein, the term "target nucleic acid" refers to a nucleic acid molecule of interest (e.g., an nucleic acid to be analyzed). In some embodiments, a target nucleic acid comprises both a target nucleotide sequence (e.g., a known or predetermined nucleotide sequence) and an adjacent nucleotide sequence which is to be determined (which may be referred to as an unknown sequence). A target nucleic acid can be of any appropriate length. In some embodiments, a target nucleic acid is double-stranded. In some embodiments, the target nucleic acid is DNA. In some embodiments, the target nucleic acid is genomic or chromosomal DNA (gDNA). In some embodiments, the target nucleic acid can be complementary DNA (cDNA). In some embodiments, the target nucleic acid is single-stranded. In some embodiments, the target nucleic acid can be RNA, e.g., mRNA, rRNA, tRNA, long non-coding RNA, microRNA.

As used herein, the term "known target nucleotide sequence" refers to a portion of a target nucleic acid for which the sequence (e.g. the identity and order of the nucleotide bases of the nucleic acid) is known. For example, in some embodiments, a known target nucleotide sequence is a nucleotide sequence of a nucleic acid that is known or that has been determined in advance of an interrogation of an adjacent unknown sequence of the nucleic acid. A known target nucleotide sequence can be of any appropriate length.

In some embodiments, a target nucleotide sequence (e.g., a known target nucleotide sequence) has a length of 10 or more nucleotides, 30 or more nucleotides, 40 or more nucleotides, 50 or more nucleotides, 100 or more nucleotides, 200 or more nucleotides, 300 or more nucleotides, 400 or more nucleotides, 500 or more nucleotides. In some embodiments, a target nucleotide sequence (e.g., a known target nucleotide sequence) has a length in range of 10 to 100 nucleotides, 10 to 500 nucleotides, 10 to 1000 nucleotides, 100 to 500 nucleotides, 100 to 1000 nucleotides, 500 to 1000 nucleotides, 500 to 5000 nucleotides.

In some embodiments, methods are provided herein for determining sequences of contiguous (or adjacent) portions

16

of a nucleic acid. As used herein, the term "nucleotide sequence contiguous to" refers to a nucleotide sequence of a nucleic acid molecule (e.g., a target nucleic acid) that is immediately upstream or downstream of another nucleotide sequence (e.g., a known nucleotide sequence). In some embodiments, a nucleotide sequence contiguous to a known target nucleotide sequence may be of any appropriate length. In some embodiments, a nucleotide sequence contiguous to a known target nucleotide sequence comprises 1 kb or less of nucleotide sequence, e.g. 1 kb or less of nucleotide sequence, 750 bp or less of nucleotide sequence, 500 bp or less of nucleotide sequence, 400 bp or less of nucleotide sequence, 300 bp or less of nucleotide sequence, 200 bp or less of nucleotide sequence, 100 bp or less of nucleotide sequence. In some embodiments, in which a sample comprises different target nucleic acids comprising a known target nucleotide sequence (e.g. a cell in which a known target nucleotide sequence occurs multiple times in its genome, or on separate, non-identical chromosomes), there may be multiple sequences which comprise "a nucleotide sequence contiguous to" the known target nucleotide sequence. As used herein, the term "determining a (or the) nucleotide sequence," refers to determining the identity and relative positions of the nucleotide bases of a nucleic acid.

In some embodiments of methods disclosed herein one or more tailed random primers are hybridized to a nucleic acid template (e.g., a template comprising a strand of a target nucleic acid). In some embodiments, a target nucleic acid is present in or obtained from a sample comprising a plurality of nucleic acids, one or more of which plurality do not comprise the target nucleic acid. In some embodiments, one or more primers (e.g., one or more tailed random primers) hybridize to substantially all of the nucleic acids in a sample. In some embodiments, one or more primers (e.g., one or more tailed random primers) hybridize to nucleic acids that comprise a target nucleic acid and to nucleic acids that do not comprise the target nucleotide sequence.

Aspects of certain methods disclosed herein relate to contacting a nucleic acid template with a plurality of different primers that share a common sequence that is 5' (or upstream) to different hybridization sequences. In some embodiments the plurality of different primers may be referred to as a population of different primers. In some embodiments, the common sequence may be referred to as a tail, as such the primers are referred to as "tailed primers." In some embodiments, different hybridization sequences of a population comprise nucleotide sequences that occur randomly or pseudorandomly within the population. In some embodiments, nucleotide sequences that occur randomly within a population contain no recognizable regularities, such that, for each nucleotide of each sequence in the population, there is an equal likelihood that the nucleotide comprises a base that is complementary with A, T, G, or C. In such embodiments, it should be appreciated that each nucleotide comprising a base that is complementary with A, T, G, or C may be a naturally occurring nucleotide, a non-naturally occurring nucleotide or a modified nucleotide.

As used herein, a "common sequence" or "shared sequence" refers to a nucleotide sequence that is present in each nucleic acid of a population of nucleic acids. In some embodiments, the common sequence is in a range of about 4 to 75, 4 to 50, 4 to 30, or 4 to 20 nucleotides in length. In some embodiments, the common sequence is 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57,

US 10,450,597 B2

17

58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, or 75 nucleotides in length.

As used herein, the term "tailed random primer" refers to a single-stranded nucleic acid molecule having a 5' nucleotide sequence (e.g., a 5' nucleotide sequence identical or complementary to a first sequencing primer) and a 3' nucleic acid sequence, in which the 3' nucleotide comprises random nucleotides (e.g., from about 3 to about 15 random nucleotides, about 6 to about 12 random nucleotides). In some embodiments, the 3' nucleotide sequence comprising random nucleotides is at least 6 nucleotides in length, e.g. 6 nucleotides or more, 7 nucleotides or more, 8 nucleotides or more, 9 nucleotides or more, 10 nucleotides or more, 11 nucleotides or more, 12 nucleotides or more, 13 nucleotides or more, 14 nucleotides or more, 15 nucleotides or more, 20 nucleotides or more, 25 nucleotides or more in length. In some embodiments, the 3' nucleotide sequence comprising random nucleotides is 3 to 6 nucleotides in length, 3 to 9 nucleotides in length, 3 to 12 nucleotides in length, 5 to 9 nucleotides in length, 6 to 12 nucleotides in length, 3 to 25 nucleotides in length, 6 to 15 nucleotides in length, or 6 to 25 nucleotides in length. In some embodiments, a tailed random primer can further comprise a spacer between the 5' nucleotide sequence and the 3' nucleotide sequence comprising about 6 to about 12 random nucleotides. In some embodiments, the spacer is a molecular barcode, e.g., that independently tags a template nucleic acid (e.g., a template RNA). In some embodiments, the spacer may be 3 to 6 nucleotides in length, 3 to 12 nucleotides in length, 3 to 25 nucleotides in length, 3 to 45 nucleotides in length, 6 to 12 nucleotides in length, 8 to 16 nucleotides in length, 6 to 25 nucleotides in length, or 6 to 45 nucleotides in length. In some embodiments, for a population of primers, the spacer is composed of random nucleotides (e.g., in which each of N is independently selected from A, G, C, and T). In some embodiments, the spacer (e.g., a molecular barcode (MBC)) is flanked by two common regions that are complementary. In some embodiments, the complementary common regions base pair to form the stem of a hairpin having a loop portion that comprises the MBC (e.g., as depicted in FIG. 3). In some embodiments, this hairpin configuration protects the MBC from annealing to other targets inhibiting the RT reaction of the extension reaction in the case of 5' fusions. In some embodiments, a population of tailed random primers can comprise individual primers with varying 3' sequences. In some embodiments, a population of tailed random primers can comprise individual primers with identical 5' nucleotide sequences, e.g., they are all compatible with the same sequencing primer. In some embodiments, a population of tailed random primers can comprise individual primers with varying 5' nucleotide sequences, e.g. an first individual primer is compatible with a first sequencing primer and a second individual primer is compatible with a second sequencing primer.

As used herein, a "hybridization sequence" refers to a sequence of a nucleic acid, such as a portion of a primer, that has sufficient complementarity with a sequence of another nucleic acid (e.g., a template molecule, a target sequence) to enable hybridization between nucleic acid. In some embodiments, the hybridization sequence is about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more nucleotides in length. In some embodiments, the hybridization sequence is in a range of 5 to 50 nucleotides in length, 5 to 40 nucleotides in length, 5 to 35 nucleotides in length, 5 to 30 nucleotides in length, 5 to 25 nucleotides in length, 5 to 20

18

nucleotides in length, 5 to 15 nucleotides in length, 5 to 10 nucleotides in length, 10 to 40 nucleotides in length, 10 to 30 nucleotides in length, or 10 to 20 nucleotides in length.

In some embodiments, methods described herein comprises an extension regimen or step. In such embodiments, extension may proceed from one or more hybridized tailed random primers, using the nucleic acid molecules which the primers are hybridized to as templates. Extension steps are described herein. In some embodiments, one or more tailed random primers can hybridize to substantially all of the nucleic acids in a sample, many of which may not comprise a known target nucleotide sequence. Accordingly, in some embodiments, extension of random primers may occur due to hybridization with templates that do not comprise a known target nucleotide sequence.

In some embodiments, methods described herein may involve a polymerase chain reaction (PCR) amplification regimen, involving one or more amplification cycles. As used herein, the term "amplification regimen" refers to a process of specifically amplifying (increasing the abundance of) a nucleic acid of interest. In some embodiments, exponential amplification occur when products of a previous polymerase extension serve as templates for successive rounds of extension. In some embodiments, a PCR amplification regimen according to methods disclosed herein may comprise at least one, and in some cases at least 5 or more iterative cycles. In some embodiments each iterative cycle comprises steps of: 1) strand separation (e.g., thermal denaturation); 2) oligonucleotide primer annealing to template molecules; and 3) nucleic acid polymerase extension of the annealed primers. It should be appreciated that any suitable conditions and times involved in each of these steps may be used. In some embodiments, conditions and times selected may depend on the length, sequence content, melting temperature, secondary structural features, or other factors relating to the nucleic acid template and/or primers used in the reaction. In some embodiments, an amplification regimen according to methods described herein is performed in a thermal cycler, many of which are commercially available.

In some embodiments, a nucleic acid extension reaction involves the use of a nucleic acid polymerase. As used herein, the phrase "nucleic acid polymerase" refers to an enzyme that catalyzes the template-dependent polymerization of nucleoside triphosphates to form primer extension products that are complementary to the template nucleic acid sequence. A nucleic acid polymerase enzyme initiates synthesis at the 3' end of an annealed primer and proceeds in the direction toward the 5' end of the template. Numerous nucleic acid polymerases are known in the art and commercially available. One group of nucleic acid polymerases are thermostable, i.e., they retain function after being subjected to temperatures sufficient to denature annealed strands of complementary nucleic acids, e.g. 94° C., or sometimes higher. A non-limiting example of a protocol for amplification involves using a polymerase (e.g., Phoenix Taq, VeraSeq) under the following conditions: 98° C. for 30 s, followed by 14-22 cycles comprising melting at 98° C. for 10 s, followed by annealing at 68° C. for 30 s, followed by extension at 72° C, 3 min, followed by holding of the reaction at 4° C. However, other appropriate reaction conditions may be used. In some embodiments, annealing/extension temperatures may be adjusted to account for differences in salt concentration (e.g., 3° C. higher to higher salt concentrations). In some embodiments, slowing the ramp rate (e.g., 1° C./s, 0.5° C./s, 0.28° C./s, 0.1° C./s or

US 10,450,597 B2

19

slower), for example, from 98° C. to 65° C., improves primer performance and coverage uniformity in highly multiplexed samples.

In some embodiments, a nucleic acid polymerase is used under conditions in which the enzyme performs a template-dependent extension. In some embodiments, the nucleic acid polymerase is DNA polymerase I, Taq polymerase, Pfu polymerase, Taq polymerase, Phusion polymerase, T4 polymerase, T7 polymerase, Klenow fragment, Klenow exo-, phi29 polymerase, AMV reverse transcriptase, M-MuLV reverse transcriptase, HIV-1 reverse transcriptase, VeraSeq Ultra polymerase, VeraSeq HF 2.0 polymerase, EnzScript or another appropriate polymerase. In some embodiments, a nucleic acid polymerase is not a reverse transcriptase. In some embodiments, a nucleic acid polymerase acts on a DNA template. In some embodiments, the nucleic acid polymerase acts on an RNA template. In some embodiments, an extension reaction involves reverse transcription performed on an RNA to produce a complementary DNA molecule (RNA-dependent DNA polymerase activity). In some embodiments, a reverse transcriptase is a mouse moloney murine leukemia virus (M-MLV) polymerase, AMV reverse transcriptase, RSV reverse transcriptase, HIV-1 reverse transcriptase, HIV-2 reverse transcriptase or another appropriate reverse transcriptase.

In some embodiments, a nucleic acid amplification reaction involves cycles including a strand separation step generally involving heating of the reaction mixture. As used herein, the term “strand separation” or “separating the strands” means treatment of a nucleic acid sample such that complementary double-stranded molecules are separated into two single strands available for annealing to an oligonucleotide primer. In some embodiments, strand separation according to methods described herein is achieved by heating the nucleic acid sample above its melting temperature (T_m). In some embodiments, for a sample containing nucleic acid molecules in a reaction preparation suitable for a nucleic acid polymerase, heating to 94° C. is sufficient to achieve strand separation. In some embodiments, a suitable reaction preparation contains one or more salts (e.g., 1 to 100 mM KCl, 0.1 to 10 MgCl₂), at least one buffering agent (e.g., 1 to 20 mM Tris-HCl), and a carrier (e.g., 0.01 to 0.5% BSA). A non-limiting example of a suitable buffer comprises 50 mM KCl, 10 mM Tris-HCl (pH 8.8@25° C.), 0.5 to 3 mM MgCl₂, and 0.1% BSA.

In some embodiments, a nucleic acid amplification involves annealing primers to nucleic acid templates having a strands characteristic of a target nucleic acid. In some embodiments, a strand of a target nucleic acid can serve as a template nucleic acid.

As used herein, the term “anneal” refers to the formation of one or more complementary base pairs between two nucleic acids. In some embodiments, annealing involve two complementary or substantially complementary nucleic acid strands hybridizing together. In some embodiments, in the context of an extension reaction annealing involves the hybridize of primer to a template such that a primer extension substrate for a template-dependent polymerase enzyme is formed. In some embodiments, conditions for annealing (e.g., between a primer and nucleic acid template) may vary based of the length and sequence of a primer. In some embodiments, conditions for annealing are based upon a T_m (e.g., a calculated T_m) of a primer. In some embodiments, an annealing step of an extension regimen involves reducing the temperature following strand separation step to a temperature based on the T_m (e.g., a calculated T_m) for a primer, for a time sufficient to permit such annealing. In some

20

embodiments, a T_m can be determined using any of a number of algorithms (e.g., OLIGO™ (Molecular Biology Insights Inc. Colorado) primer design software and VENTRO NTI™ (Invitrogen, Inc. California) primer design software and programs available on the internet, including Primer3, Oligo Calculator, and NetPrimer (Premier Biosoft, Palo Alto, Calif.; and freely available on the world wide web (e.g., at premierbiosoft.com/netprimer/netprlaunch/Help/xnetpr-launch.html). In some embodiments, the T_m of a primer can be calculated using following formula, which is used by NetPrimer software and is described in more detail in Frieir et al. PNAS 1986 83:9373-9377 which is incorporated by reference herein in its entirety.

$$T_m = \frac{\Delta H}{R \ln(C/4)} + 16.6 \log([K^+]/(1+0.7[K^+])) - \frac{273.15}{273.15}$$

wherein, ΔH is enthalpy for helix formation; ΔS is entropy for helix formation; R is molar gas constant (1.987 cal/° C.*mol); C is the nucleic acid concentration; and $[K^+]$ is salt concentration. For most amplification regimens, the annealing temperature is selected to be about 5° C. below the predicted T_m , although temperatures closer to and above the T_m (e.g., between 1° C. and 5° C. below the predicted T_m or between 1° C. and 5° C. above the predicted T_m) can be used, as can, for example, temperatures more than 5° C. below the predicted T_m (e.g., 6° C. below, 8° C. below, 10° C. below or lower). In some embodiments, the closer an annealing temperature is to the T_m , the more specific is the annealing. In some embodiments, the time used for primer annealing during an extension reaction (e.g., within the context of a PCR amplification regimen) is determined based, at least in part, upon the volume of the reaction (e.g., with larger volumes involving longer times). In some embodiments, the time used for primer annealing during an extension reaction (e.g., within the context of a PCR amplification regimen) is determined based, at least in part, upon primer and template concentrations (e.g., with higher relative concentrations of primer to template involving less time than lower relative concentrations). In some embodiments, depending upon volume and relative primer/template concentration, primer annealing steps in an extension reaction (e.g., within the context of an amplification regimen) can be in the range of 1 second to 5 minutes, 10 seconds and 2 minutes, or 30 seconds to 2 minutes. As used herein, “substantially anneal” refers to an extent to which complementary base pairs form between two nucleic acids that, when used in the context of a PCR amplification regimen, is sufficient to produce a detectable level of a specifically amplified product.

As used herein, the term “polymerase extension” refers to template-dependent addition of at least one complementary nucleotide, by a nucleic acid polymerase, to the 3' end of a primer that is anneal to a nucleic acid template. In some embodiments, polymerase extension adds more than one nucleotide, e.g., up to and including nucleotides corresponding to the full length of the template. In some embodiments, conditions for polymerase extension are based, at least in part, on the identity of the polymerase used. In some embodiments, the temperature used for polymerase extension is based upon the known activity properties of the enzyme. In some embodiments, in which annealing temperatures are below the optimal temperatures for the enzyme, it may be acceptable to use a lower extension temperature. In some embodiments, enzymes may retain at least partial activity below their optimal extension temperatures. In some embodiments, a polymerase extension (e.g., performed thermostable polymerases) (e.g., Taq polymerase and variants

US 10,450,597 B2

21

thereof) is performed at 65° C. to 75° C. or 68° C. to 72° C. In some embodiments, methods provided herein involve polymerase extension of primers that are anneal to nucleic acid templates at each cycle of a PCR amplification regimen. In some embodiments, a polymerase extension is performed using a polymerase that has relatively strong strand displacement activity. In some embodiments, polymerases having strong strand displacement are useful for preparing nucleic acids for purposes of detecting fusions (e.g., 5' fusions).

In some embodiments, primer extension is performed under conditions that permit the extension of annealed oligonucleotide primers. As used herein, the term "conditions that permit the extension of an annealed oligonucleotide such that extension products are generated" refers to the set of conditions including, for example temperature, salt and co-factor concentrations, pH, and enzyme concentration under which a nucleic acid polymerase catalyzes primer extension. In some embodiments, such conditions are based, at least in part, on the nucleic acid polymerase being used. In some embodiments, a polymerase may perform a primer extension reaction in a suitable reaction preparation. In some embodiments, a suitable reaction preparation contains one or more salts (e.g., 1 to 100 mM KCl, 0.1 to 10 MgCl₂), at least one buffering agent (e.g., 1 to 20 mM Tris-HCl), a carrier (e.g., 0.01 to 0.5% BSA) and one or more NTPs (e.g., 10 to 200 μM of each of dATP, dTTP, dCTP, and dGTP). non-limiting set of conditions is 50 mM KCl, 10 mM Tris-HCl (pH 8.8@25° C.), 0.5 to 3 mM MgCl₂, 200 μM each dNTP, and 0.1% BSA at 72° C., under which a polymerase (e.g., Taq polymerase) catalyzes primer extension. In some embodiments, conditions for initiation and extension may include the presence of one, two, three or four different deoxyribonucleoside triphosphates (e.g., selected from dATP, dTTP, dCTP, and dGTP) and a polymerization-inducing agent such as DNA polymerase or reverse transcriptase, in a suitable buffer. In some embodiments, a "buffer" may include solvents (e.g., aqueous solvents) plus appropriate cofactors and reagents which affect pH, ionic strength, etc.).

In some embodiments, nucleic acid amplification involve up to 5, up to 10, up to 20, up to 30, up to 40 or more rounds (cycles) of amplification. In some embodiments, nucleic acid amplification may comprise a set of cycles of a PCR amplification regimen from 5 cycles to 20 cycles in length. In some embodiments, an amplification step may comprise a set of cycles of a PCR amplification regimen from 10 cycles to 20 cycles in length. In some embodiments, each amplification step can comprise a set of cycles of a PCR amplification regimen from 12 cycles to 16 cycles in length. In some embodiments, an annealing temperature can be less than 70° C. In some embodiments, an annealing temperature can be less than 72° C. In some embodiments, an annealing temperature can be about 65° C. In some embodiments, an annealing temperature can be from about 61 to about 72° C.

In various embodiments, methods and compositions described herein relate to performing a PCR amplification regimen with one or more of the types of primers described herein. As used herein, "primer" refers to an oligonucleotide capable of specifically annealing to a nucleic acid template and providing a 3' end that serves as a substrate for a template-dependent polymerase to produce an extension product which is complementary to the template. In some embodiments, a primer useful in methods described herein is single-stranded, such that the primer and its complement can anneal to form two strands. Primers according to methods and compositions described herein may comprise a hybridization sequence (e.g., a sequence that anneals with a nucleic acid template) that is less than or equal to 300

22

nucleotides in length, e.g., less than or equal to 300, or 250, or 200, or 150, or 100, or 90, or 80, or 70, or 60, or 50, or 40, or 30 or fewer, or 20 or fewer, or 15 or fewer, but at least 6 nucleotides in length. In some embodiments, a hybridization sequence of a primer may be 6 to 50 nucleotides in length, 6 to 35 nucleotides in length, 6 to 20 nucleotides in length, 10 to 25 nucleotides in length.

Any suitable method may be used for synthesizing oligonucleotides and primers. In some embodiments, commercial sources offer oligonucleotide synthesis services suitable for providing primers for use in methods and compositions described herein, e.g. INVITROGEN™ Custom DNA Oligos; Life Technologies; Grand Island, N.Y. or custom DNA Oligos from IDT; Coralville, Iowa).

In some embodiments, after an extension from a tailed random primer has occurred, the extension product and template can be amplified in a first amplification step. In some embodiments, amplification may involve a set of PCR amplification cycles using a first target-specific primer and a first tail primer. In some embodiments, the amplification may result in at least part of the tailed random primer sequence present in the extension product being amplified. In some embodiments, the amplification may result in all of the tailed random primer sequence present in the extension product being amplified.

As used herein, the term "first target-specific primer" refers to a single-stranded oligonucleotide comprising a nucleic acid sequence that can specifically anneal under suitable annealing conditions to a nucleic acid template that has a strand characteristic of a target nucleic acid.

In some embodiments, a primer (e.g., a target specific primer) can comprise a 5' tag sequence portion. In some embodiments, multiple primers (e.g., all first-target specific primers) present in a reaction can comprise identical 5' tag sequence portions. In some embodiments, in a multiplex PCR reaction, different primer species can interact with each other in an off-target manner, leading to primer extension and subsequently amplification by DNA polymerase. In such embodiments, these primer dimers tend to be short, and their efficient amplification can overtake the reaction and dominate resulting in poor amplification of desired target sequence. Accordingly, in some embodiments, the inclusion of a 5' tag sequence in primers (e.g., on target specific primer(s)) may result in formation of primer dimers that contain the same complementary tails on both ends. In some embodiments, in subsequent amplification cycles, such primer dimers would denature into single-stranded DNA primer dimers, each comprising complementary sequences on their two ends which are introduced by the 5' tag. In some embodiments, instead of primer annealing to these single stranded DNA primer dimers, an intra-molecular hairpin (a panhandle like structure) formation may occur due to the proximate accessibility of the complementary tags on the same primer dimer molecule instead of an inter-molecular interaction with new primers on separate molecules. Accordingly, in some embodiments, these primer dimers may be inefficiently amplified, such that primers are not exponentially consumed by the dimers for amplification; rather the tagged primers can remain in high and sufficient concentration for desired specific amplification of target sequences. In some embodiments, accumulation of primer dimers may be undesirable in the context of multiplex amplification because they compete for and consume other reagents in the reaction.

In some embodiments, a 5' tag sequence can be a GC-rich sequence. In some embodiments, a 5' tag sequence may comprise at least 50% GC content, at least 55% GC content,

US 10,450,597 B2

23

at least 60% GC content, at least 65% GC content, at least 70% GC content, at least 75% GC content, at least 80% GC content, or higher GC content. In some embodiments, a tag sequence may comprise at least 60% GC content. In some embodiments, a tag sequence may comprise at least 65% GC content.

In some embodiments, a target-specific primer (e.g., a second target-specific primer) is a single-stranded oligonucleotide comprising a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of a known target nucleotide sequence of an amplicon of an amplification reaction, and a 5' portion comprising a tag sequence (e.g., a nucleotide sequence that is identical to or complementary to a sequencing primer (e.g., a second sequencing primer)).

In some embodiments, a second target-specific primer of an amplification regimen is nested with respect to a first target-specific primer of the amplification regimen. In some embodiments, the second target-specific primer is nested with respect to the first target-specific primer by at least 3 nucleotides, e.g., by 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, or 15 or more nucleotides. In some embodiments, all of the target-specific primers (e.g., second target-specific primers) used in an amplification regimen comprise the same 5' portion. In some embodiments, the 5' portion target-specific primer can be configured to suppress primer dimers as described herein.

In some embodiments, first and second target-specific primers are used in an amplification regimen that are substantially complementary to the same strand of a target nucleic acid. In some embodiments, portions of the first and second target-specific primers that specifically anneal to a target sequence (e.g., a known target sequence) can comprise a total of at least 20 unique bases of the known target nucleotide sequence, e.g., 20 or more unique bases, 25 or more unique bases, 30 or more unique bases, 35 or more unique bases, 40 or more unique bases, or 50 or more unique bases. In some embodiments, portions of first and second target-specific primers that specifically anneal to a target sequence (e.g., a known target sequence) can comprise a total of at least 30 unique bases of the known target nucleotide sequence.

As used herein, the term "first tail primer" refers to a nucleic acid molecule comprising a nucleic acid sequence identical to the tail portion of tailed primer.

As used herein, the term "second tail primer" refers to a nucleic acid molecule comprising a nucleic acid sequence identical to a portion of a first sequencing primer, adapter, index primer, etc. and is optionally nested with respect to a first tailed primer. In some embodiments, the second tail primer sits outside of the first tail primer to facilitate addition of appropriate index tags, adapters (e.g., for use in a sequencing platform), etc. In some embodiments, a second tailed primer is identical to a sequencing primer. In some embodiments, a second tailed primer is complementary to a sequencing primer.

In some embodiments, a second tail primer is nested with respect to a first tail primer. In some embodiments, a second tail primer is not nested with respect to a first tail primer. In some embodiments, tail primers of an amplification regimen are nested with respect to one another by at least 3 nucleotides, e.g., by 3 nucleotides, by 4 nucleotides, by 5 nucleotides, by 6 nucleotides, by 7 nucleotides, by 8 nucleotides, by 9 nucleotides, by 10 nucleotides or more.

In some embodiments, a first tail primer comprises a nucleic acid sequence identical to or complementary to the extension product of step (b) strand which is not comprised

24

by the second tail primer and which is located closer to the 5' end of the tailed random primer than any of the sequence identical to or complementary to the second tail primer. Thus, in some embodiments, a second tail primer sits outside of a region added by a random tail primer (5' end), e.g., within the 5' tail added by the first tail primers.

In some embodiments, a first tail primer can comprise a nucleic acid sequence identical to or complementary to a stretch (e.g., of about 20 nucleotides) of the 5'-most nucleotides of a tailed random primer, and a second tail primer can comprise a nucleic acid sequence identical to or complementary to about 30 bases of a tailed random primer, with a 5' nucleotides that is at least 3 nucleotides 3' of the 5' terminus of the tailed random primer.

In some embodiments, use of nested tail primers minimizes or eliminates the production of final amplicons that are amplifiable (e.g., during bridge PCR or emulsion PCR) but cannot be sequenced, a situation that can arise during hemi-nested methods. In some embodiments, hemi-nested approaches using a primer identical to a sequencing primer can result in the carry-over of undesired amplification products from a first PCR step to a second PCR step and may yield artificial sequencing reads. In some embodiments, the use of two tail primers, as described herein can reduce, and in some embodiments eliminate, these problems.

In some embodiments, in a first PCR amplification cycle of a first amplification step, a first target-specific primer can specifically anneal to a template strand of any nucleic acid comprising the known target nucleotide sequence. In some embodiments, depending upon the orientation with which the first target-specific primer was designed, sequence upstream or downstream of the known target nucleotide sequence, and complementary to the template strand will be synthesized. In some embodiments, in which an extension product is formed that comprises the hybridization sequence with which the first target-specific primer forms complementary base pairs, a double-stranded amplification product can be formed that comprises the first target-specific primer (and the sequence complementary thereto), the target nucleotide sequence downstream of the first target-specific primer (and the sequence complementary thereto), and the tailed random primer sequence (and the sequence complementary thereto). In such embodiments, in subsequent PCR amplification cycles, both the first target-specific primer and the first tail primer are capable of specifically annealing to appropriate strands of the amplification product and the sequence between the known nucleotide target sequence and the tailed random primer can be amplified.

In some embodiments, of methods described herein, a portion of an amplified product (an amplicon) is amplified in further rounds of amplification. In some embodiments, the further rounds of amplification may involve PCR amplification cycles performed using a second target-specific primer and a first sequencing primer or a second tail primer. In some embodiments, a PCR amplification cycles may involve the use of PCR parameters identical to, or which differ from, those of one or more other (e.g., prior) of PCR amplification cycles. In some embodiments, PCR amplification regimens can have the same or different annealing temperatures or the same or different extension step time lengths.

In some embodiments, methods described herein allow for determining the nucleotide sequence contiguous to a known target nucleotide sequence on either or both flanking regions of the known target nucleotide sequence. Regardless of whether the target nucleic acid normally exists as a single-stranded or double-stranded nucleic acid, sequence

US 10,450,597 B2

25

information may be represented in a single-stranded format (Strand A), from 5' to 3'. In some embodiments, if the sequence 5' to a known target nucleotide sequence of Strand A is to be determined, gene-specific primers can be complementary to (anneal to) Strand A. If the sequence 3' to a known target nucleotide sequence of Strand A is to be determined, the gene-specific primers can be identical to Strand A, such that they will anneal to the complementary strand of a double-stranded target nucleic acid.

In some embodiments, methods described herein, relating to the use of a first and second gene-specific primer can result in assays with a superior on-target rate, e.g. 70-90%. In some embodiments, the assays and methods described herein can have a target specificity rate of at least 85%.

In some embodiments, primers disclosed herein (e.g., target-specific primers, tail primers) are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of from about 61 to 72° C., e.g. from about 61 to 69° C., from about 63 to 69° C., from about 63 to 67° C., from about 64 to 66° C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of less than 72° C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of less than 70° C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of less than 68° C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of about 65° C.

In some embodiments, portions of the target-specific primers that specifically anneal to the known target nucleotide sequence will anneal specifically at a temperature of about 61 to 72° C., e.g. from about 61 to 69° C., from about 63 to 69° C., from about 63 to 67° C., from about 64 to 66° C. In some embodiments, portions of the target-specific primers that specifically anneal to the known target nucleotide sequence will anneal specifically at a temperature of about 65° C. in a PCR buffer.

In some embodiments, primers described herein do not comprise modified bases (e.g. the primers can not comprise a blocking 3' amine). However, in some embodiments, primers described herein do comprise modified or non-naturally occurring bases. In some embodiments, primers may be modified with a label capable of providing a detectable signal, either directly or indirectly. Non-limiting examples of such labels include radioisotopes, fluorescent molecules, biotin, and others. In some embodiments, primers disclosed herein may include contain a biotin linker or other suitable linker (e.g., for conjugating the primer to a support). In some embodiments, primer may contain a target sequence of an endonucleases such that cleavage with the appropriate enzyme. In other embodiments, the 5' end of a primer may include a sequence that is complementary with a nucleic acid bound to a bead or other support, e.g., a flow cell substrate. Primers may or may not comprise modified internucleoside linkages.

In some embodiments, of methods described herein, nucleic acids (e.g., amplified nucleic acids, extension products, target nucleic acids) can be sequenced. In some embodiments, sequencing can be performed by a next-generation sequencing method. As used herein "next-generation sequencing" refers to oligonucleotide sequencing technologies that have the capacity to sequence oligonucle-

26

otides at speeds above those possible with conventional sequencing methods (e.g. Sanger sequencing), due to performing and reading out thousands to millions of sequencing reactions in parallel. Non-limiting examples of next-generation sequencing methods/platforms include Massively Parallel Signature Sequencing (Lynx Therapeutics); 454 pyrosequencing (454 Life Sciences/Roche Diagnostics); solid-phase, reversible dye-terminator sequencing (Solexa/Illumina); SOLiD technology (Applied Biosystems); Ion semiconductor sequencing (ION Torrent); DNA nanoball sequencing (Complete Genomics); and technologies available from Pacific Biosciences, Intelligen Bio-systems, Oxford Nanopore Technologies, and Helicos Biosciences. In some embodiments, the sequencing primers can comprise portions compatible with the selected next-generation sequencing method. Next-generation sequencing technologies and the constraints and design parameters of associated sequencing primers are well known in the art (see, e.g. Shendure, et al., "Next-generation DNA sequencing," *Nature*, 2008, vol. 26, No. 10, 1135-1145; Mardis, "The impact of next-generation sequencing technology on genetics," *Trends in Genetics*, 2007, vol. 24, No. 3, pp. 133-141; Su, et al., "Next-generation sequencing and its applications in molecular diagnostics" *Expert Rev Mol Diagn*, 2011, 11(3):333-43; Zhang et al., "The impact of next-generation sequencing on genomics", *J Genet Genomics*, 2011, 38(3): 95-109; (Nyren, P. et al. *Anal Biochem* 208: 17175 (1993); Bentley, D. R. *Curr Opin Genet Dev* 16:545-52 (2006); Strausberg, R. L., et al. *Drug Disc Today* 13:569-77 (2008); U.S. Pat. No. 7,282,337; 7,279,563; 7,226,720; 7,220,549; 7,169,560; 6,818,395; 6,911,345; US Pub. Nos. 2006/0252077; 2007/0070349; and 2007/0070349; which are incorporated by reference herein in their entireties).

In some embodiments, the sequencing step involve the use of a first and second sequencing primers. In some embodiments, the first and second sequencing primers are selected to be compatible with a next-generation sequencing method as described herein.

Methods of aligning sequencing reads to known sequence databases of genomic and/or cDNA sequences are well known in the art and software is commercially available for this process. In some embodiments, reads (less the sequencing primer nucleotide sequence) which do not map, in their entirety, to wild-type sequence databases can be genomic rearrangements or large indel mutations. In some embodiments, reads (less the sequencing primer nucleotide sequence) comprising sequences which map to multiple locations in the genome can be genomic rearrangements.

In some embodiments, primers may contain additional sequences such as an identifier sequence (e.g., a barcode, an index), sequencing primer hybridization sequences (e.g., Rd1), and adapter sequences. In some embodiments the adapter sequences are sequences used with a next generation sequencing system. In some embodiments, the adapter sequences are P5 and P7 sequences for Illumina-based sequencing technology. In some embodiments, the adapter sequence are P1 and A compatible with Ion Torrent sequencing technology.

In some embodiments, as used herein, a "barcode," "molecular barcode," "molecular barcode tag" and "index" may be used interchangeably, generally referring to a nucleotide sequence of a nucleic acid that is useful as an identifier, such as, for example, a source identifier, location identifier, date or time identifier (e.g., date or time of sampling or processing), or other identifier of the nucleic acid. In some embodiments, such barcode or index sequences are useful for identifying different aspects of a

US 10,450,597 B2

27

nucleic acid that is present in a population of nucleic acids. In some embodiments, barcode or index sequences may provide a source or location identifier for a target nucleic acid. For example, a barcode or index sequence may serve to identify a patient from whom a nucleic acid is obtained. In some embodiments, barcode or index sequences enable sequencing of multiple different samples on a single reaction (e.g., performed in a single flow cell). In some embodiments, an index sequence can be used to orientate a sequence imager for purposes of detecting individual sequencing reactions. In some embodiments, a barcode or index sequence may be 2 to 25 nucleotides in length, 2 to 15 nucleotides in length, 2 to 10 nucleotides in length, 2 to 6 nucleotides in length. In some embodiments, a barcode or index comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or at least 25 nucleotides.

In some embodiments, when a population of tailed random primers is used in accordance with methods described herein, multiple distinguishable amplification products can be present after amplification. In some embodiments, because tailed random primers hybridize at various positions throughout nucleic acid molecules of a sample, a set of target-specific primers can hybridize (and amplify) the extension products created by more than 1 hybridization event, e.g. one tailed random primer may hybridize at a first distance (e.g., 100 nucleotides) from a target-specific primer hybridization site, and another tailed random primer can hybridize at a second distance (e.g., 200 nucleotides) from a target-specific primer hybridization site, thereby resulting in two amplification products (e.g., a ~100 bp amplification product and a ~200 bp amplification product). In some embodiments, these multiple amplification products can each be sequenced in. In some embodiments, sequencing of these multiple amplification products is advantageous because it provides multiple overlapping sequence reads that can be compared with one another to detect sequence errors introduced during amplification or sequencing processes. In some embodiments, individual amplification products can be aligned and where they differ in the sequence present at a particular base, an artifact or error of PCR and/or sequencing may be present.

In some embodiments, target nucleic acids and/or amplification products thereof can be isolated from enzymes, primers, or buffer components before and/or after any of appropriate step of a method. Any suitable methods for isolating nucleic acids may be used. In some embodiments, the isolation can comprise Solid Phase Reversible Immobilization (SPRI) cleanup. Methods for SPRI cleanup are well known in the art and kits are commercially available, e.g. Agencourt AMPure XP-PCR Purification (Cat No. A63880, Beckman Coulter, Brea, Calif.). In some embodiments, enzymes can be inactivated by heat treatment.

In some embodiments, unhybridized primers can be removed from a nucleic acid preparation using appropriate methods (e.g., purification, digestion, etc.). In some embodiments, a nuclease (e.g., exonuclease I) is used to remove primer from a preparation. In some embodiments, such nucleases are heat inactivated subsequent to primer digestion. Once the nucleases are inactivated a further set of primers may be added together with other appropriate components (e.g., enzymes, buffers) to perform a further amplification reaction.

In some embodiments, a target nucleic acid genomic DNA or a portion thereof. In some embodiments, a target nucleic acid can be ribonucleic acid (RNA), e.g. mRNA, or

28

a portion thereof. In some embodiments, a target nucleic acid can be a cDNA or a portion thereof.

Many of the sequencing methods suitable for use in methods described herein provide sequencing runs with optimal read lengths of tens to hundreds of nucleotide bases (e.g. Ion Torrent technology can produce read lengths of 200-400 bp). Target nucleic acids may or may not be substantially longer than this optimal read length. In some embodiments, in order for an amplified nucleic acid portion to be of a suitable length for use in a particular sequencing technology, the average distance between the known target nucleotide sequence and an end of the target nucleic acid to which a tailed random primer is hybridizable should be as close to the optimal read length of the selected technology as possible. In some embodiments, if the optimal read-length of a given sequencing technology is 200 bp, then the nucleic acid molecules amplified in accordance with methods described herein should have an average length of about 800 bp, about 700 bp, about 600 bp, about 500 bp, about 400 bp, about 300 bp, about 200 bp or less.

Nucleic acids used herein (e.g., prior to sequencing) can be sheared, e.g. mechanically or enzymatically sheared, to generate fragments of any desired size. Non-limiting examples of mechanical shearing processes include sonication, nebulization, and AFA™ shearing technology available from Covaris (Woburn, Mass.). In some embodiments, a nucleic acid can be mechanically sheared by sonication.

In some embodiments, a target nucleic acid is not sheared or digested. In some embodiments, nucleic acid products of preparative steps (e.g., extension products, amplification products) are not sheared or enzymatically digested.

In some embodiments, when a target nucleic acid is an RNA, the sample can be subjected to a reverse transcriptase regimen to generate DNA template and the DNA template can then be sheared. In some embodiments, target RNA can be sheared before performing a reverse transcriptase regimen. In some embodiments, a sample comprising target RNA can be used in methods described herein using total nucleic acids extracted from either fresh or degraded specimens; without the need of genomic DNA removal for cDNA sequencing; without the need of ribosomal RNA depletion for cDNA sequencing; without the need of mechanical or enzymatic shearing in any of the steps; by subjecting the RNA for double-stranded cDNA synthesis using random hexamers.

In some embodiments, a known target nucleic acid can contain a fusion sequence resulting from a gene rearrangement. In some embodiments, methods described herein are suited for determining the presence and/or identity of a gene rearrangement. In some embodiments, identity of one portion of a gene rearrangement is previously known (e.g., the portion of a gene rearrangement that is to be targeted by the gene-specific primers) and the sequence of the other portion may be determined using methods disclosed herein. In some embodiments, a gene rearrangement can involve an oncogene. In some embodiments, a gene rearrangement can comprise a fusion oncogene.

In some embodiments, a target nucleic acid is present in or obtained from an appropriate sample (e.g., a food sample, environmental sample, biological sample e.g., blood sample, etc.). In some embodiments, the sample is a biological sample obtained from a subject. In some embodiments a sample can be a diagnostic sample obtained from a subject. In some embodiments, a sample can further comprise proteins, cells, fluids, biological fluids, preservatives, and/or other substances. By way of non-limiting example, a sample can be a cheek swab, blood, serum, plasma, sputum, cerebrospinal

US 10,450,597 B2

29

fluid, urine, tears, alveolar isolates, pleural fluid, pericardial fluid, cyst fluid, tumor tissue, tissue, a biopsy, saliva, an aspirate, or combinations thereof. In some embodiments, a sample can be obtained by resection or biopsy.

In some embodiments, the sample can be obtained from a subject in need of treatment for a disease associated with a genetic alteration, e.g. cancer or a hereditary disease. In some embodiments, a known target sequence is present in a disease-associated gene.

In some embodiments, a sample is obtained from a subject in need of treatment for cancer. In some embodiments, the sample comprises a population of tumor cells, e.g. at least one tumor cell. In some embodiments, the sample comprises a tumor biopsy, including but not limited to, untreated biopsy tissue or treated biopsy tissue (e.g. formalin-fixed and/or paraffin-embedded biopsy tissue).

In some embodiments, the sample is freshly collected. In some embodiments, the sample is stored prior to being used in methods and compositions described herein. In some embodiments, the sample is an untreated sample. As used herein, "untreated sample" refers to a biological sample that has not had any prior sample pre-treatment except for dilution and/or suspension in a solution. In some embodiments, a sample is obtained from a subject and preserved or processed prior to being utilized in methods and compositions described herein. By way of non-limiting example, a sample can be embedded in paraffin wax, refrigerated, or frozen. A frozen sample can be thawed before determining the presence of a nucleic acid according to methods and compositions described herein. In some embodiments, the sample can be a processed or treated sample. Exemplary methods for treating or processing a sample include, but are not limited to, centrifugation, filtration, sonication, homogenization, heating, freezing and thawing, contacting with a preservative (e.g. anti-coagulant or nuclease inhibitor) and any combination thereof. In some embodiments, a sample can be treated with a chemical and/or biological reagent. Chemical and/or biological reagents can be employed to protect and/or maintain the stability of the sample or nucleic acid comprised by the sample during processing and/or storage. In addition, or alternatively, chemical and/or biological reagents can be employed to release nucleic acids from other components of the sample. By way of non-limiting example, a blood sample can be treated with an anti-coagulant prior to being utilized in methods and compositions described herein. Suitable methods and processes for processing, preservation, or treatment of samples for nucleic acid analysis may be used in the method disclosed herein. In some embodiments, a sample can be a clarified fluid sample, for example, by centrifugation. In some embodiments, a sample can be clarified by low-speed centrifugation (e.g. 3,000xg or less) and collection of the supernatant comprising the clarified fluid sample.

In some embodiments, a nucleic acid present in a sample can be isolated, enriched, or purified prior to being utilized in methods and compositions described herein. Suitable methods of isolating, enriching, or purifying nucleic acids from a sample may be used. For example, kits for isolation of genomic DNA from various sample types are commercially available (e.g. Catalog Nos. 51104, 51304, 56504, and 56404; Qiagen; Germantown, Md.). In some embodiments, methods described herein relate to methods of enriching for target nucleic acids, e.g., prior to a sequencing of the target nucleic acids. In some embodiments, a sequence of one end of the target nucleic acid to be enriched is not known prior to sequencing. In some embodiments, methods described herein relate to methods of enriching specific nucleotide

30

sequences prior to determining the nucleotide sequence using a next-generation sequencing technology. In some embodiments, methods of enriching specific nucleotide sequences do not comprise hybridization enrichment.

Methods described herein can be employed in a multiplex format. In embodiments of methods described herein, multiplex applications can include determining the nucleotide sequence contiguous to one or more known target nucleotide sequences. As used herein, "multiplex amplification" refers to a process involve simultaneous amplification of more than one target nucleic acid in one reaction vessel. In some embodiments, methods involve subsequent determination of the sequence of the multiplex amplification products using one or more sets of primers. Multiplex can refer to the detection of between about 2-1,000 different target sequences in a single reaction. As used herein, multiplex refers to the detection of any range between 2-1,000, e.g., between 5-500, 25-1000, or 10-100 different target sequences in a single reaction, etc. The term "multiplex" as applied to PCR implies that there are primers specific for at least two different target sequences in the same PCR reaction.

In some embodiments, target nucleic acids in a sample, or separate portions of a sample, can be amplified with a plurality of primers (e.g., a plurality of first and second target-specific primers). In some embodiments, the plurality of primers (e.g., a plurality of first and second target-specific primers) can be present in a single reaction mixture, e.g. multiple amplification products can be produced in the same reaction mixture. In some embodiments, the plurality of primers (e.g., a plurality of sets of first and second target-specific primers) can specifically anneal to known target sequences comprised by separate genes. In some embodiments, at least two sets of primers (e.g., at least two sets of first and second target-specific primers) can specifically anneal to different portions of a known target sequence. In some embodiments, at least two sets of primers (e.g., at least two sets of first and second target-specific primers) can specifically anneal to different portions of a known target sequence comprised by a single gene. In some embodiments, at least two sets of primers (e.g., at least two sets of first and second target-specific primers) can specifically anneal to different exons of a gene comprising a known target sequence. In some embodiments, the plurality of primers (e.g., first target-specific primers) can comprise identical 5' tag sequence portions.

In embodiments of methods described herein, multiplex applications can include determining the nucleotide sequence contiguous to one or more known target nucleotide sequences in multiple samples in one sequencing reaction or sequencing run. In some embodiments, multiple samples can be of different origins, e.g. from different tissues and/or different subjects. In such embodiments, primers (e.g., tailed random primers) can further comprise a barcode portion. In some embodiments, a primer (e.g., a tailed random primer) with a unique barcode portion can be added to each sample and ligated to the nucleic acids therein; the samples can subsequently be pooled. In such embodiments, each resulting sequencing read of an amplification product will comprise a barcode that identifies the sample containing the template nucleic acid from which the amplification product is derived.

In some embodiments of methods described herein, a determination of the sequence contiguous to a known oligonucleotide target sequence can provide information relevant to treatment of disease. Thus, in some embodiments, methods disclosed herein can be used to aid in treating

US 10,450,597 B2

31

disease. In some embodiments, a sample can be from a subject in need of treatment for a disease associated with a genetic alteration. In some embodiments, a known target sequence a sequence of a disease-associated gene, e.g. an oncogene. In some embodiments, a sequence contiguous to a known oligonucleotide target sequence and/or the known oligonucleotide target sequence can comprise a mutation or genetic abnormality which is disease-associated, e.g. a SNP, an insertion, a deletion, and/or a gene rearrangement. In some embodiments, a sequence contiguous to a known target sequence and/or a known target sequence present in a sample comprised sequence of a gene rearrangement product. In some embodiments, a gene rearrangement can be an oncogene, e.g. a fusion oncogene.

Certain treatments for cancer are particularly effective against tumors comprising certain oncogenes, e.g. a treatment agent which targets the action or expression of a given fusion oncogene can be effective against tumors comprising that fusion oncogene but not against tumors lacking the fusion oncogene. Methods described herein can facilitate a determination of specific sequences that reveal oncogene status (e.g. mutations, SNPs, and/or rearrangements). In some embodiments, methods described herein can further allow the determination of specific sequences when the sequence of a flanking region is known, e.g. methods described herein can determine the presence and identity of gene rearrangements involving known genes (e.g., oncogenes) in which the precise location and/or rearrangement partner are not known before methods described herein are performed.

In some embodiments, technology described herein relates to a method of treating cancer. Accordingly, in some embodiments, methods provided herein may involve detecting, in a tumor sample obtained from a subject in need of treatment for cancer, the presence of one or more oncogene rearrangements; and administering a cancer treatment which is effective against tumors having any of the detected oncogene rearrangements. In some embodiments, technology described herein relates to a method of determining if a subject in need of treatment for cancer will be responsive to a given treatment. Accordingly, in some embodiments, methods provided herein may involve detecting, in a tumor sample obtained from a subject, the presence of an oncogene rearrangement, in which the subject is determined to be responsive to a treatment targeting an oncogene rearrangement product if the presence of the oncogene rearrangement is detected.

In some embodiments, a subject is in need of treatment for lung cancer. In some embodiments, e.g. when the sample is obtained from a subject in need of treatment for lung cancer, the known target sequence can comprise a sequence from a gene selected from the group of ALK, ROS1, and RET. Accordingly, in some embodiments, gene rearrangements result in fusions involving the ALK, ROS1, or RET. Non-limiting examples of gene arrangements involving ALK, ROS1, or RET are described in, e.g., Soda et al. *Nature* 2007 448:561-6; Rikova et al. *Cell* 2007 131:1190-1203; Kohno et al. *Nature Medicine* 2012 18:375-7; Takouchi et al. *Nature Medicine* 2012 18:378-81; which are incorporated by reference herein in their entirety. However, it should be appreciated that the precise location of a gene rearrangement, and the identity of the second gene involved in the rearrangement may not be known in advance. Accordingly, in methods described herein, the presence and identity of such rearrangements can be detected without having to know the location of the rearrangement or the identity of the second gene involved in the gene rearrangement.

32

In some embodiments, the known target sequence can comprise sequence from a gene selected from the group of: ALK, ROS1, and RET.

In some embodiments, the presence of a gene rearrangement of ALK in a sample obtained from a tumor in a subject can indicate that the tumor is susceptible to treatment with a treatment selected from the group consisting of: an ALK inhibitor; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; diamino and aminopyrimidine inhibitors of ALK kinase activity such as NVP-TAE684 and PF-02341066 (see, e.g. Galkin et al. *Proc Natl Acad Sci USA*, 2007, 104:270-275; Zou et al. *Cancer Res.* 2007, 67:4408-4417; Hallberg and Palmer *F1000 Med Reports* 2011 3:21; and Sakamoto et al. *Cancer Cell* 2011 19:679-690) and molecules disclosed in WO 04/079326. All of the foregoing references are incorporated by reference herein in their entirety. An ALK inhibitor can include any agent that reduces the expression and/or kinase activity of ALK or a portion thereof, including, e.g. oligonucleotides, small molecules, and/or peptides that reduce the expression and/or activity of ALK or a portion thereof. As used herein "anaplastic lymphoma kinase" or "ALK" refers to a transmembrane tyrosine kinase typically involved in neuronal regulation in the wildtype form. The nucleotide sequence of the ALK gene and mRNA are known for a number of species, including human (e.g. SEQ ID NO: 2 (mRNA), NCBI Gene ID: 238).

In some embodiments, the presence of a gene rearrangement of ROS1 in a sample obtained from a tumor in a subject can indicate that the tumor is susceptible to treatment with a treatment selected from the group consisting of: a ROS1 inhibitor and an ALK inhibitor as described herein above (e.g. crizotinib). A ROS1 inhibitor can include any agent that reduces the expression and/or kinase activity of ROS1 or a portion thereof, including, e.g. oligonucleotides, small molecules, and/or peptides that reduce the expression and/or activity of ROS1 or a portion thereof. As used herein "c-ros oncogene 1" or "ROS1" (also referred to in the art as *ros-1*) refers to a transmembrane tyrosine kinase of the sevenless subfamily and which interacts with PTPN6. Nucleotide sequences of the ROS1 gene and mRNA are known for a number of species, including human (e.g. SEQ ID NO: 1 (mRNA), NCBI Gene ID: 238).

In some embodiments, the presence of a gene rearrangement of RET in a sample obtained from a tumor in a subject can indicate that the tumor is susceptible to treatment with a treatment selected from the group consisting of: a RET inhibitor; DP-2490, DP-3636, SU5416; BAY 43-9006, BAY 73-4506 (regorafenib), ZD6474, NVP-AST487, sorafenib, RPI-1, XL184, vandetanib, sunitinib, imatinib, pazopanib, axitinib, motesanib, gefitinib, and withaferin A (see, e.g. Samadi et al. *Surgery* 2010 148:1228-36; Cuccuru et al. *JNCI* 2004 13:1006-1014; Akeno-Stuart et al. *Cancer Research* 2007 67:6956; Grazma et al. *J Clin Oncol* 2010 28:15 s 5559; Mologni et al. *J Mol Endocrinol* 2006 37:199-212; Calmomagno et al. *Journal NCI* 2006 98:326-334; Mologni. *Curr Med Chem* 2011 18:162-175 and the compounds disclosed in WO 06/034833; US Patent Publication 2011/0201598 and U.S. Pat. No. 8,067,434). All of the foregoing references are incorporated by reference herein in their entirety. A RET inhibitor can include any agent that reduces the expression and/or kinase activity of RET or a portion thereof, including, e.g. oligonucleotides, small molecules, and/or peptides that reduce the expression and/or activity of RET or a portion thereof. As used herein "rearranged during transfection" or "RET" refers to a receptor

US 10,450,597 B2

33

tyrosine kinase of the cad herein superfamily which is involved in neural crest development and recognizes glial cell line-derived neurotrophic factor family signaling molecules. Nucleotide sequences of the RET gene and mRNA are known for a number of species, including human (e.g. SEQ ID NOs: 3-4 (mRNA), NCBI Gene ID: 5979).

Further non-limiting examples of applications of methods described herein include detection of hematological malignancy markers and panels thereof (e.g. including those to detect genomic rearrangements in lymphomas and leukemias), detection of sarcoma-related genomic rearrangements and panels thereof; and detection of IGH/TCR gene rearrangements and panels thereof for lymphoma testing.

In some embodiments, methods described herein relate to treating a subject having or diagnosed as having, e.g. cancer with a treatment for cancer. Subjects having cancer can be identified by a physician using current methods of diagnosing cancer. For example, symptoms and/or complications of lung cancer which characterize these conditions and aid in diagnosis are well known in the art and include but are not limited to, weak breathing, swollen lymph nodes above the collarbone, abnormal sounds in the lungs, dullness when the chest is tapped, and chest pain. Tests that may aid in a diagnosis of, e.g. lung cancer include, but are not limited to, x-rays, blood tests for high levels of certain substances (e.g. calcium), CT scans, and tumor biopsy. A family history of lung cancer, or exposure to risk factors for lung cancer (e.g. smoking or exposure to smoke and/or air pollution) can also aid in determining if a subject is likely to have lung cancer or in making a diagnosis of lung cancer.

Cancer can include, but is not limited to, carcinoma, including adenocarcinoma, lymphoma, blastoma, melanoma, sarcoma, leukemia, squamous cell cancer, small-cell lung cancer, non-small cell lung cancer, gastrointestinal cancer, Hodgkin's and non Hodgkin's lymphoma, pancreatic cancer, glioblastoma, basal cell carcinoma, biliary tract cancer, bladder cancer, brain cancer including glioblastomas and medulloblastomas; breast cancer, cervical cancer, choriocarcinoma; colon cancer, colorectal cancer, endometrial carcinoma, endometrial cancer; esophageal cancer, gastric cancer; various types of head and neck cancers, intraepithelial neoplasms including Bowen's disease and Paget's disease; hematological neoplasms including acute lymphocytic and myelogenous leukemia; Kaposi's sarcoma, hairy cell leukemia; chronic myelogenous leukemia, AIDS-associated leukemias and adult T-cell leukemia lymphoma; kidney cancer such as renal cell carcinoma, T-cell acute lymphoblastic leukemia/lymphoma, lymphomas including Hodgkin's disease and lymphocytic lymphomas; liver cancer such as hepatic carcinoma and hepatoma, Merkel cell carcinoma, melanoma, multiple myeloma; neuroblastomas; oral cancer including squamous cell carcinoma; ovarian cancer including those arising from epithelial cells, sarcomas including leiomyosarcoma, rhabdomyosarcoma, liposarcoma, fibrosarcoma, and osteosarcoma; pancreatic cancer; skin cancer including melanoma, stromal cells, germ cells and mesenchymal cells; prostate cancer, rectal cancer; vulval cancer, renal cancer including adenocarcinoma; testicular cancer including germinal tumors such as seminoma, non-seminoma (teratomas, choriocarcinomas), stromal tumors, and germ cell tumors; thyroid cancer including thyroid adenocarcinoma and medullary carcinoma; esophageal cancer, salivary gland carcinoma, and Wilms' tumors. In some embodiments, the cancer can be lung cancer.

In some embodiments, methods described herein comprise administering an effective amount of compositions described herein, e.g. a treatment for cancer to a subject in

34

order to alleviate a symptom of a cancer. As used herein, "alleviating a symptom of a cancer" is ameliorating any condition or symptom associated with the cancer. As compared with an equivalent untreated control, such reduction is by at least 5%, 10%, 20%, 40%, 50%, 60%, 80%, 90%, 95%, 99% or more as measured by any standard technique. A variety of means for administering the compositions described herein to subjects are known to those of skill in the art. Such methods can include, but are not limited to oral, parenteral, intravenous, intramuscular, subcutaneous, transdermal, airway (aerosol), pulmonary, cutaneous, topical, injection, or intratumoral administration. Administration can be local or systemic. The term "effective amount" as used herein refers to the amount of a treatment needed to alleviate at least one or more symptom of the disease or disorder, and relates to a sufficient amount of pharmacological composition to provide the desired effect. The term "therapeutically effective amount" therefore refers to an amount that is sufficient to effect a particular anti-cancer effect when administered to a typical subject. An effective amount as used herein, in various contexts, would also include an amount sufficient to delay the development of a symptom of the disease, alter the course of a symptom disease (for example but not limited to, slowing the progression of a symptom of the disease), or reverse a symptom of the disease. Thus, it is not generally practicable to specify an exact "effective amount". However, for any given case, an appropriate "effective amount" can be determined by one of ordinary skill in the art using only routine experimentation. The effects of any particular dosage can be monitored by a suitable bioassay. The dosage can be determined by a physician and adjusted, as appropriate, to suit observed effects of the treatment.

Non-limiting examples of a treatment for cancer can include radiation therapy, surgery, gemcitabine, cisplatin, paclitaxel, carboplatin, bortezomib, AMG479, vorinostat, rituximab, temozolomide, rapamycin, ABT-737, PI-103; alkylating agents such as thiopeta and CYTOXAN® cyclophosphamide; alkyl sulfonates such as busulfan, improsulfan and piposulfan; aziridines such as benzodopa, carboquone, meturedopa, and uredopa; ethylenimines and methylamelamines including altretamine, triethylenemelamine, triethylenephosphoramide, triethylenethiophosphoramide and trimethylolmelamine; acetogenins (especially bullatacin and bullatacinone); a camptothecin (including the synthetic analogue topotecan); bryostatins; calystatin; CC-1065 (including its adozelesin, carzelesin and bizelesin synthetic analogues); cryptophycins (particularly cryptophycin 1 and cryptophycin 8); dolastatin; duocarmycin (including the synthetic analogues, KW-2189 and CB1-TM1); eleutherobin; pancratistatin; a sarcodictyin; spongistatin; nitrogen mustards such as chlorambucil, chlornaphazine, cholophosphamide, estramustine, ifosfamide, mechlorethamine, mechlorethamine oxide hydrochloride, melphalan, novembichin, phenesterine, prednimustine, trofosfamide, uracil mustard; nitrosoureas such as carmustine, chlorozotocin, fotemustine, lomustine, nimustine, and ranimustine; antibiotics such as the enediyne antibiotics (e.g., calicheamicin, especially calicheamicin gamma1 and calicheamicin omega1 (see, e.g., Agnew, Chem. Intl. Ed. Engl., 33: 183-186 (1994)); dynemicin, including dynemicin A; bisphosphonates, such as clodronate; an esperamicin; as well as neocarzinostatin chromophore and related chromoprotein enediyne antibiotic chromophores), aclacinomysins, actinomycin, anthramycin, azaserine, bleomycins, cactinomycin, carubicin, caminomycin, carzinophilin, chromomycin, dactinomycin, daunorubicin, detorubicin, 6-di-

azo-5-oxo-L-norleucine, ADRIAMYCIN® doxorubicin (including morpholino-doxorubicin, cyanomorpholino-doxorubicin, 2-pyrrolino-doxorubicin and deoxydoxorubicin), epirubicin, esorubicin, idarubicin, marcellomycin, mitomycins such as mitomycin C, mycophenolic acid, nogalamycin, olivomycins, peplomycin, potfiromycin, puromycin, que-lamycin, rodorubicin, streptonigrin, streptozocin, tubercidin, ubenimex, zinoastatin, zorubicin; anti-metabolites such as methotrexate and 5-fluorouracil (5-FU); folic acid analogues such as denopterin, methotrexate, pteropterin, trimetrexate; purine analogs such as fludarabine, 6-mercaptopurine, thia-miprine, thioguanine; pyrimidine analogs such as ancit-abine, azacitidine, 6-azauridine, carmofur, cytarabine, dide-oxuridine, doxifluridine, enocitabine, floxuridine; androgens such as calusterone, dromostanolone propionate, epitostanol, mepitostane, testolactone; anti-adrenals such as aminoglutethimide, mitotane, trilostane; folic acid replen-isher such as frolinic acid; aceglutone; aldophosphamide glycoside; aminolevulinic acid; eniluracil; amsacrine; bestabucil; bisantrene; edatraxate; defofamine; demecol-cine; diaziquone; elformithine; elliptinium acetate; an epoth-ilone; etoglucid; gallium nitrate; hydroxyurea; lentinan; lonidamine; maytansinoids such as maytansine and ansami-tocins; mitoguanzone; mitoxantrone; mopidanmol; nitraerine; pentostatin; phenamet; pirarubicin; losoxantrone; podoph-yllic acid; 2-ethylhydrazide; procarbazine; PSK® poly-saccharide complex (JHS Natural Products, Eugene, Ore.); razoxane; rhizoxin; sizofuran; spirogermanium; tenuazonic acid; triaziquone; 2,2',2"-trichlorotriethylamine; trichoth-ecenes (especially T-2 toxin, verracurin A, roridin A and anguidine); urethan; vindesine; dacarbazine; mannomustine; mitobronitol; mitolactol; pipobroman; gacytosine; arabino-side ("Ara-C"); cyclophosphamide; thioplea; taxoids, e.g., TAXOL® paclitaxel (Bristol-Myers Squibb Oncology, Princeton, N.J.), ABRAXANE® Cremophor-free, albumin-engineered nanoparticle formulation of paclitaxel (Ameri-can Pharmaceutical Partners, Schaumburg, Ill.), and TAXO-TERE® doxetaxel (Rhone-Poulenc Rorer, Antony, France); chloranbucil; GEMZAR® gemcitabine; 6-thioguanine; mer-captopurine; methotrexate; platinum analogs such as cispla-tin, oxaliplatin and carboplatin; vinblastine; platinum; etoposide (VP-16); ifosfamide; mitoxantrone; vincristine; NAVELBINE® vinorelbine; novantrone; teniposide; eda-trexate; daunomycin; aminopterin; xeloda; ibandronate; iri-notecan (Camptosar, CPT-11) (including the treatment regi-men of irinotecan with 5-FU and leucovorin); topoisomerase inhibitor RFS 2000; difluoromethylornithine (DMFO); ret-inoids such as retinoic acid; capecitabine; combretastatin; leucovorin (LV); oxaliplatin, including the oxaliplatin treat-ment regimen (FOLFOX); lapatinib (Tykerb®); inhibitors of PKC- α , Raf, H-Ras, EGFR (e.g., erlotinib (Tarceva®)) and VEGF-A that reduce cell proliferation and pharmaceu-tically acceptable salts, acids or derivatives of any of the above. In addition, methods of treatment can further include the use of radiation or radiation therapy. Further, methods of treatment can further include the use of surgical treatments.

In some embodiments, methods described herein can be applicable for resequencing, e.g. for confirming particularly relevant, low-quality, and/or complex sequences obtained by non-directed sequencing of a large amount of nucleic acids. By way of non-limiting examples, methods described herein can allow the directed and/or targeted resequencing of targeted disease gene panels (e.g. 10-100 genes), resequencing to confirm variants obtained in large scale sequencing projects, whole exome resequencing, and/or targeted resequencing for detection of single nucleotide variants, mul-

tipule nucleotide variants, insertions, deletions, copy number changes, and methylation status.

In some embodiments, methods described herein can allow microbiota sequencing, ancient sample sequencing, and/or new variant virus genotyping.

For convenience, the meaning of some terms and phrases used in the specification, examples, and appended claims, are provided below. Unless stated otherwise, or implicit from context, the following terms and phrases include the meanings provided below. The definitions are provided to aid in describing particular embodiments, and are not intended to limit the claimed invention, because the scope of the invention is limited only by the claims. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. If there is an apparent discrepancy between the usage of a term in the art and its definition provided herein, the definition provided within the specification shall prevail.

For convenience, certain terms employed herein, in the specification, examples and appended claims are collected here.

The terms "decrease", "reduced", "reduction", or "inhibit" are all used herein generally to mean a decrease by a statistically significant amount. However, for avoidance of doubt, "reduced", "reduction", "decrease", or "inhibit" means a decrease by at least 10% as compared to a reference level, for example a decrease by at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and including a 100% decrease (e.g. absent level or non-detectable level as compared to a reference level), or any decrease between 10-100% as compared to a reference level. In the context of a marker or symptom is meant a statistically significant decrease in such level. The decrease can be, for example, at least 10%, at least 20%, at least 30%, at least 40% or more, and is preferably down to a level accepted as within the range of normal for an individual without such disorder.

The terms "increased", "increase", "enhance", or "activate" are all used herein to generally mean an increase by a statically significant amount; for the avoidance of doubt, the terms "increased", "increase", "enhance", or "activate" mean an increase of at least 10% as compared to a reference level, for example an increase of at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and including a 100% increase or any increase between 10-100% as compared to a reference level, or at least about a 2-fold, or at least about a 3-fold, or at least about a 4-fold, or at least about a 5-fold or at least about a 10-fold increase, or any increase between 2-fold and 10-fold or greater as compared to a reference level.

As used herein, a "subject" means a human or animal. Usually the animal is a vertebrate such as a primate, rodent, domestic animal or game animal. Primates include chimpanzees, cynomolgous monkeys, spider monkeys, and macaques, e.g., Rhesus. Rodents include mice, rats, wood-chucks, ferrets, rabbits and hamsters. Domestic and game animals include cows, horses, pigs, deer, bison, buffalo, feline species, e.g., domestic cat, canine species, e.g., dog, fox, wolf, avian species, e.g., chicken, emu, ostrich, and fish, e.g., trout, catfish and salmon. In some embodiments, the subject is a mammal, e.g., a primate, e.g., a human. The terms, "individual," "patient" and "subject" are used interchangeably herein.

US 10,450,597 B2

37

Preferably, the subject is a mammal. The mammal can be a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but is not limited to these examples. Mammals other than humans can be advantageously used as subjects that represent animal models of, e.g. lung cancer. A subject can be male or female.

A subject can be one who has been previously diagnosed with or identified as suffering from or having a condition in need of treatment (e.g. cancer) or one or more complications related to such a condition, and optionally, have already undergone treatment for the condition or the one or more complications related to the condition. Alternatively, a subject can also be one who has not been previously diagnosed as having the condition (e.g. cancer) or one or more complications related to the condition. For example, a subject can be one who exhibits one or more risk factors for the condition or one or more complications related to the condition or a subject who does not exhibit risk factors.

A "subject in need" of treatment for a particular condition can be a subject having that condition, diagnosed as having that condition, or at risk of developing that condition.

As used herein, a "disease associated with a genetic alteration" refers to any disease which is caused by, at least in part, by an alteration in the genetic material of the subject as compared to a healthy wildtype subject, e.g. a deletion, an insertion, a SNP, a gene rearrangement. A disease can be caused by, at least in part, an alteration in the genetic material of the subject if the alteration increases the risk of the subject developing the disease, increases the subject's susceptibility to a disease (including infectious diseases, or diseases with an infectious component), causes the production of a disease-associated molecule, or causes cells to become diseased or abnormal (e.g. loss of cell cycle regulation in cancer cells). Diseases can be associated with multiple genetic alterations, e.g. cancers.

As used herein, the term "nucleic acid" refers to any molecule, preferably a polymeric molecule, incorporating units of ribonucleic acid, deoxyribonucleic acid or an analog thereof. The nucleic acid can be either single-stranded or double-stranded. A single-stranded nucleic acid can be one strand nucleic acid of a denatured double-stranded DNA. Alternatively, it can be a single-stranded nucleic acid not derived from any double-stranded DNA. In one aspect, the template nucleic acid is DNA. In another aspect, the template is RNA. Suitable nucleic acid molecules are DNA, including genomic DNA or cDNA. Other suitable nucleic acid molecules are RNA, including mRNA.

The term "isolated" or "partially purified" as used herein refers, in the case of a nucleic acid, to a nucleic acid separated from at least one other component (e.g., nucleic acid or polypeptide) that is present with the nucleic acid as found in its natural source and/or that would be present with the nucleic acid when expressed by a cell. A chemically synthesized nucleic acid or one synthesized using in vitro transcription/translation is considered "isolated."

As used herein, the term "complementary" refers to the ability of nucleotides to form hydrogen-bonded base pairs. In some embodiment, complementary refers to hydrogen-bonded base pair formation preferences between the nucleotide bases G, A, T, C and U, such that when two given polynucleotides or polynucleotide sequences anneal to each other, A pairs with T and G pairs with C in DNA, and G pairs with C and A pairs with U in RNA. As used herein, "substantially complementary" refers to a nucleic acid molecule or portion thereof (e.g. a primer) having at least 90% complementarity over the entire length of the molecule or portion thereof with a second nucleotide sequence, e.g. 90%

38

complementary, 95% complementary, 98% complementary, 99% complementary, or 100% complementary. As used herein, "substantially identical" refers to a nucleic acid molecule or portion thereof having at least 90% identity over the entire length of a the molecule or portion thereof with a second nucleotide sequence, e.g. 90% identity, 95% identity, 98% identity, 99% identity, or 100% identity.

As used herein, "specific" when used in the context of a primer specific for a target nucleic acid refers to a level of complementarity between the primer and the target such that there exists an annealing temperature at which the primer will anneal to and mediate amplification of the target nucleic acid and will not anneal to or mediate amplification of non-target sequences present in a sample.

As used herein, "amplified product", "amplification product", or "amplicon" refers to oligonucleotides resulting from an amplification reaction that are copies of a portion of a particular target nucleic acid template strand and/or its complementary sequence, which correspond in nucleotide sequence to the template nucleic acid sequence and/or its complementary sequence. An amplification product can further comprise sequence specific to the primers and which flanks sequence which is a portion of the target nucleic acid and/or its complement. An amplified product, as described herein will generally be double-stranded DNA, although reference can be made to individual strands thereof.

As used herein, a "portion" of a nucleic acid molecule refers to contiguous set of nucleotides comprised by that molecule. A portion can comprise all or only a subset of the nucleotides comprised by the molecule. A portion can be double-stranded or single-stranded.

As used herein, the terms "treat," "treatment," "treating," or "amelioration" refer to therapeutic treatments, wherein the object is to reverse, alleviate, ameliorate, inhibit, slow down or stop the progression or severity of a condition associated with a disease or disorder, e.g. lung cancer. The term "treating" includes reducing or alleviating at least one adverse effect or symptom of a condition, disease or disorder associated with a condition. Treatment is generally "effective" if one or more symptoms or clinical markers are reduced. Alternatively, treatment is "effective" if the progression of a disease is reduced or halted. That is, "treatment" includes not just the improvement of symptoms or markers, but also a cessation of, or at least slowing of, progress or worsening of symptoms compared to what would be expected in the absence of treatment. Beneficial or desired clinical results include, but are not limited to, alleviation of one or more symptom(s), diminishment of extent of disease, stabilized (i.e., not worsening) state of disease, delay or slowing of disease progression, amelioration or palliation of the disease state, remission (whether partial or total), and/or decreased mortality, whether detectable or undetectable. The term "treatment" of a disease also includes providing relief from the symptoms or side-effects of the disease (including palliative treatment).

The term "statistically significant" or "significantly" refers to statistical significance and generally means a two standard deviation (2SD) below normal, or lower, concentration of the marker.

Other than in the operating examples, or where otherwise indicated, all numbers expressing quantities of ingredients or reaction conditions used herein should be understood as modified in all instances by the term "about." The term "about" when used in connection with percentages can mean $\pm 1\%$.

As used herein the term "comprising" or "comprises" is used in reference to compositions, methods, and respective

US 10,450,597 B2

39

component(s) thereof, that are essential to method or composition, yet open to the inclusion of unspecified elements, whether essential or not.

The term "consisting of" refers to compositions, methods, and respective components thereof as described herein, which are exclusive of any element not recited in that description of the embodiment.

As used herein the term "consisting essentially of" refers to those elements required for a given embodiment. The term permits the presence of elements that do not materially affect the basic and novel or functional characteristic(s) of that embodiment.

The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of this disclosure, suitable methods and materials are described below. The abbreviation, "e.g." is derived from the Latin *exempli gratia*, and is used herein to indicate a non-limiting example. Thus, the abbreviation "e.g." is synonymous with the term "for example."

Definitions of common terms in cell biology and molecular biology can be found in "The Merck Manual of Diagnosis and Therapy", 19th Edition, published by Merck Research Laboratories, 2006 (ISBN 0-911910-19-0); Robert S. Porter et al. (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9). Definitions of common terms in molecular biology can also be found in Benjamin Lewin, Genes X, published by Jones & Bartlett Publishing, 2009 (ISBN-10: 0763766321); Kendrew et al. (eds.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8) and Current Protocols in Protein Sciences 2009, Wiley Intersciences, Coligan et al., eds.

Unless otherwise stated, the present invention was performed using standard procedures, as described, for example in Sambrook et al., Molecular Cloning: A Laboratory Manual (3 ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., USA (2001); and Davis et al., Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York, USA (1995) which are all incorporated by reference herein in their entireties.

Other terms are defined herein within the description of the various aspects of the invention.

All patents and other publications, including literature references, issued patents, published patent applications, and co-pending patent applications, cited throughout this application are expressly incorporated herein by reference for the purpose of describing and disclosing, for example, methodologies described in such publications that might be used in connection with technology described herein. These publications are provided solely for their disclosure prior to the filing date of the present application. Nothing in this regard should be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention or for any other reason. All statements as to the date or representation as to the contents of these documents is based on the information available to the applicants and does not constitute any admission as to the correctness of the dates or contents of these documents.

The description of embodiments of the disclosure is not intended to be exhaustive or to limit the disclosure to the precise form disclosed. While specific embodiments of, and examples for, the disclosure are described herein for illus-

40

trative purposes, various equivalent modifications are possible within the scope of the disclosure, as those skilled in the relevant art will recognize. For example, while method steps or functions are presented in a given order, alternative embodiments may perform functions in a different order, or functions may be performed substantially concurrently. The teachings of the disclosure provided herein can be applied to other procedures or methods as appropriate. The various embodiments described herein can be combined to provide further embodiments. Aspects of the disclosure can be modified, if appropriate, to employ the compositions, functions and concepts of the above references and application to provide yet further embodiments of the disclosure. These and other changes can be made to the disclosure in light of the detailed description. All such modifications are intended to be included within the scope of the appended claims.

Specific elements of any of the foregoing embodiments can be combined or substituted for elements in other embodiments. Furthermore, while advantages associated with certain embodiments of the disclosure have been described in the context of these embodiments, other embodiments may also exhibit such advantages, and not all embodiments need necessarily exhibit such advantages to fall within the scope of the disclosure.

Technology described herein is further illustrated by the following examples which in no way should be construed as being further limiting.

Some embodiments of the technology described herein can be defined according to any of the following numbered paragraphs:

1. A method of determining the nucleotide sequence contiguous to a known target nucleotide sequence, the method comprising:
 - (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with an initial target-specific primer under hybridization conditions;
 - (b) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid molecule as a template;
 - (c) contacting the product of step (b) with a population of tailed random primers under hybridization conditions;
 - (d) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template;
 - (e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer;
 - (f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer;
 - (g) sequencing the amplified portion from step (f) using a first and second sequencing primer;
- wherein the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides;
- wherein the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the

US 10,450,597 B2

41

known target nucleotide sequence of the target nucleic acid at the annealing temperature;

wherein the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer;

wherein the first tail primer comprises a nucleic acid sequence identical to the tailed random primer; and

wherein the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

2. A method of determining the nucleotide sequence contiguous to a known target nucleotide sequence, the method comprising:

(a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions;

(b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template;

(c) contacting the product of step (b) with an initial target-specific primer under hybridization conditions;

(d) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid molecule as a template;

(e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer;

(f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer;

(g) sequencing the amplified portion from step (f) using a first and second sequencing primer;

wherein the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides;

wherein the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature;

wherein the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer;

wherein the first tail primer comprises a nucleic acid sequence identical to the tailed random primer; and

wherein the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

42

3. The method of any of paragraphs 1-2, further comprising a step of contacting the sample and products with RNase after extension of the initial target-specific primer.

4. The method of any of paragraphs 1-3, wherein the tailed random primer can form a hair-pin loop structure.

5. The method of any of paragraphs 1-4, wherein the initial target-specific primer and the first target-specific primer are identical.

6. The method of any of paragraphs 1-5, wherein the tailed random primer further comprises a barcode portion comprising 6-12 random nucleotides between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising 6-12 random nucleotides.

7. A method of determining the nucleotide sequence contiguous to a known target nucleotide sequence, the method comprising:

(a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions;

(b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template;

(c) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer;

(d) amplifying a portion of the amplicon resulting from step (c) with a second tail primer and a second target-specific primer;

(e) sequencing the amplified portion from step (d) using a first and second sequencing primer;

wherein the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer; a middle barcode portion comprising; and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides;

wherein the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature;

wherein the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer;

wherein the first tail primer comprises a nucleic acid sequence identical to the tailed random primer; and

wherein the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

8. The method of paragraph 7, wherein the each tailed random primer further comprises a spacer nucleic acid sequence between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising about 6 to about 12 random nucleotides.

US 10,450,597 B2

43

9. The method of paragraph 7 or 8, wherein the unhybridized primers are removed from the reaction after an extension step.
10. The method of any of paragraphs 7-9, wherein the second tail primer is nested with respect to the first tail primer by at least 3 nucleotides.
11. The method of any of paragraphs 7-10, wherein the first target-specific primer further comprises a 5' tag sequence portion comprising a nucleic acid sequence of high GC content which is not substantially complementary to or substantially identical to any other portion of any of the primers.
12. The method of any of paragraphs 7-11, wherein the second tail primer is identical to the full-length first sequencing primer.
13. The method of any of paragraphs 7-12, wherein the portions of the target-specific primers that specifically anneal to the known target will anneal specifically at a temperature of about 65° C. in a PCR buffer.
14. The method of any of paragraphs 7-13, wherein the sample comprises genomic DNA.
15. The method of any of paragraphs 7-14, wherein the sample comprises RNA and the method further comprises a first step of subjecting the sample to a reverse transcriptase regimen.
16. The method of any of paragraphs 7-15, wherein the nucleic acids present in the sample have not been subjected to shearing or digestion.
17. The method of any of paragraphs 7-16, wherein the sample comprises single-stranded gDNA or cDNA.
18. The method of any of paragraphs 7-17, wherein the reverse transcriptase regimen comprises the use of random hexamers.
19. The method of any of paragraphs 7-18, wherein a gene rearrangement comprises the known target sequence.
20. The method of paragraph 19, wherein the gene rearrangement is present in a nucleic acid selected from the group consisting of: genomic DNA; RNA; and cDNA.
21. The method of any of paragraphs 19-20, wherein the gene rearrangement comprises an oncogene.
22. The method of paragraph 21, wherein the gene rearrangement comprises a fusion oncogene.
23. The method of any of paragraphs 7-22, wherein the nucleic acid product is sequenced by a next-generation sequencing method.
24. The method of paragraph 23, wherein the next-generation sequencing method comprises a method selected from the group consisting of: Ion Torrent, Illumina, SOLiD, 454; Massively Parallel Signature Sequencing solid-phase, reversible dye-terminator sequencing; and DNA nanoball sequencing.
25. The method of any of paragraphs 7-24, wherein the first and second sequencing primers are compatible with the selected next-generation sequencing method.
26. The method of any of paragraphs 7-25, wherein the method comprises contacting the sample, or separate portions of the sample, with a plurality of sets of first and second target-specific primers.
27. The method of any of paragraphs 7-26, wherein the method comprises contacting a single reaction mixture comprising the sample with a plurality of sets of first and second target-specific primers.
28. The method of any of paragraphs 7-27, wherein the plurality of sets of first and second target-specific primers specifically anneal to known target nucleotide sequences comprised by separate genes.

44

29. The method of any of paragraphs 7-28, wherein at least two sets of first and second target-specific primers specifically anneal to different portions of a known target nucleotide sequence.
30. The method of any of paragraphs 7-29, wherein at least two sets of first and second target-specific primers specifically anneal to different portions of a single gene comprising a known target nucleotide sequence.
31. The method of any of paragraphs 7-30, wherein at least two sets of first and second target-specific primers specifically anneal to different exons of a gene comprising a known nucleotide target sequence.
32. The method of any of paragraphs 7-31, wherein the plurality of first target-specific primers comprise identical 5' tag sequence portions.
33. The method of any of paragraphs 7-32, wherein each tailed random primer in a population of tailed random primers further comprises an identical sample barcoding portion.
34. The method of paragraph 33, wherein multiple samples are each contacted with a separate population of tailed random primers with a sample barcoding portion; wherein each population of tailed random primers has a distinct sample barcoding portion; and wherein the samples are pooled after step (b).
35. The method of any of paragraphs 7-34, wherein each amplification step comprises a set of cycles of a PCR amplification regimen from 5 cycles to 20 cycles in length.
36. The method of any of paragraphs 7-35, wherein the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of from about 61 to 72° C.
37. The method of any of paragraphs 7-36, wherein the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of about 65° C.
38. The method of any of paragraphs 7-37, wherein the target nucleic acid molecule is from a sample, optionally which is a biological sample obtained from a subject.
39. The method of paragraph 38, wherein the sample is obtained from a subject in need of treatment for a disease associated with a genetic alteration.
40. The method of paragraph 39, wherein the disease is cancer.
41. The method of paragraph 38, wherein the sample comprises a population of tumor cells.
42. The method of paragraph 38, wherein the sample is a tumor biopsy.
43. The method paragraph 40, wherein the cancer is lung cancer.
44. The method of any of paragraphs 7-43, wherein a disease-associated gene comprises the known target sequence.
45. The method of 38, wherein a gene rearrangement product in the sample comprises the known target sequence.
46. The method of paragraph 45, wherein the gene rearrangement product is an oncogene.
47. A method of preparing nucleic acids for analysis, the method comprising:
 - (a) contacting a nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific

US 10,450,597 B2

45

hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer; and

(b) contacting a nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers,

wherein an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers.

48. The method of paragraph 47, wherein the target nucleic acid is a ribonucleic acid.

49. The method of paragraph 47, wherein the target nucleic acid is a deoxyribonucleic acid.

50. The method of any of paragraphs 47 to 49 wherein steps (a) and (b) are performed sequentially.

51. The method of any of paragraphs 47 to 50, wherein the nucleic acid template in step (a) comprises an extension product resulting from the hybridization and extension of the at least one of the plurality of different primers in step (b).

52. The method of any one of paragraphs 47 to 50, wherein the nucleic acid template in step (b) comprises an extension product resulting from the hybridization and extension of the target-specific primer in step (a).

53. The method of paragraph 48, wherein the target nucleic acid is a messenger RNA encoded from a chromosomal segment that comprises a genetic rearrangement.

54. The method of paragraph 49, wherein the target nucleic acid is a chromosomal segment that comprises a portion of a genetic rearrangement.

55. The method of paragraph 8, wherein the genetic rearrangement is an inversion, deletion, or translocation.

56. The method of any one of paragraphs 47 to 55 further comprising amplifying the extension product.

57. The method of any one of paragraphs 47 to 55 further comprising contacting the extension product or amplified extension product with an immobilized oligonucleotide under conditions in which hybridization occurs between the extension product and immobilized oligonucleotide.

58. The method of any preceding paragraph wherein the target nucleic acid comprises a target portion having a known sequence and a flanking portion having an unknown sequence.

59. The method of paragraph 58, wherein different hybridization sequences are complementary to the flanking portion.

60. The method of paragraph 58 or 59, wherein the target-specific hybridization sequence is complementary to the target portion.

61. The method of any of paragraphs 47 to 60, wherein the target-specific primer further comprises, 5' to the target-specific hybridization sequence, at least one of an index sequence, a barcode sequence and an adaptor sequence.

62. The method of any of paragraphs 47 to 60, wherein the common sequence comprises at least one of an index sequence, barcode sequence and an adaptor sequence.

46

63. The method of any of paragraphs 1-62, wherein the adaptor sequence is a cleavable adaptor sequence for immobilizing oligonucleotides in a flow cell.

Some embodiments of the technology described herein can be defined according to any of the following numbered paragraphs:

1. A method of preparing nucleic acids for analysis, the method comprising:

(a) contacting a nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer; and

(b) contacting a nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers,

wherein an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers.

2. The method of paragraph 1, wherein the target nucleic acid is a ribonucleic acid.

3. The method of paragraph 1, wherein the target nucleic acid is a deoxyribonucleic acid.

4. The method of any of paragraphs 1 to 3 wherein steps (a) and (b) are performed sequentially.

5. The method of any of paragraphs 1 to 4, wherein the nucleic acid template in step (a) comprises an extension product resulting from the hybridization and extension of the at least one of the plurality of different primers in step (b).

6. The method of any one of paragraphs 1 to 4, wherein the nucleic acid template in step (b) comprises an extension product resulting from the hybridization and extension of the target-specific primer in step (a).

7. The method of paragraph 2, wherein the target nucleic acid is a messenger RNA encoded from a chromosomal segment that comprises a genetic rearrangement.

8. The method of paragraph 3, wherein the target nucleic acid is a chromosomal segment that comprises a portion of a genetic rearrangement.

9. The method of paragraph 8, wherein the genetic rearrangement is an inversion, deletion, or translocation.

10. The method of any one of paragraphs 1 to 9 further comprising amplifying the extension product.

11. The method of any one of paragraphs 1 to 9 further comprising contacting the extension product or amplified extension product with an immobilized oligonucleotide under conditions in which hybridization occurs between the extension product and immobilized oligonucleotide.

12. The method of any preceding paragraph wherein the target nucleic acid comprises a target portion having a known sequence and a flanking portion having an unknown sequence.

13. The method of paragraph 12, wherein different hybridization sequences are complementary to the flanking portion.

14. The method of paragraph 12 or 13, wherein the target-specific hybridization sequence is complementary to the target portion.

US 10,450,597 B2

47

15. The method of any of paragraphs 1 to 14, wherein the target-specific primer further comprises, 5' to the target-specific hybridization sequence, at least one of an index sequence, a barcode sequence and an adaptor sequence.
16. The method of any of paragraphs 1 to 14, wherein the common sequence comprises at least one of an index sequence, barcode sequence and an adaptor sequence.
17. The method of paragraph 15 or 16, wherein the adaptor sequence is a cleavable adaptor sequence for immobilizing oligonucleotides in a flow cell.
18. A method of determining the nucleotide sequence contiguous to a known target nucleotide sequence, the method comprising:
- (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with an initial target-specific primer under hybridization conditions;
 - (b) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid molecule as a template;
 - (c) contacting the product of step (b) with a population of tailed random primers under hybridization conditions;
 - (d) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template;
 - (e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer;
 - (f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer;
 - (g) sequencing the amplified portion from step (f) using a first and second sequencing primer;
- wherein the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides;
- wherein the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature;
- wherein the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (e), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer;
- wherein the first tail primer comprises a nucleic acid sequence identical to the tailed random primer; and
- wherein the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.
19. A method of determining the nucleotide sequence contiguous to a known target nucleotide sequence, the method comprising:

48

- (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions;
 - (b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template;
 - (c) contacting the product of step (b) with an initial target-specific primer under hybridization conditions;
 - (d) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid molecule as a template;
 - (e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer;
 - (f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer;
 - (g) sequencing the amplified portion from step (f) using a first and second sequencing primer;
- wherein the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides;
- wherein the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature;
- wherein the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer;
- wherein the first tail primer comprises a nucleic acid sequence identical to the tailed random primer; and
- wherein the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.
20. The method of any of paragraphs 18-19, further comprising a step of contacting the sample and products with RNase after extension of the initial target-specific primer.
21. The method of any of paragraphs 1-3, wherein the tailed random primer can form a hair-pin loop structure.
22. The method of any of paragraphs 1-4, wherein the initial target-specific primer and the first target-specific primer are identical.
23. The method of any of paragraphs 1-5, wherein the tailed random primer further comprises a barcode portion comprising 6-12 random nucleotides between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising 6-12 random nucleotides. 7. A method of determining the nucleotide sequence contiguous to a known target nucleotide sequence, the method comprising:

US 10,450,597 B2

49

- (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions;
 - (b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template;
 - (c) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer;
 - (d) amplifying a portion of the amplicon resulting from step (c) with a second tail primer and a second target-specific primer;
 - (e) sequencing the amplified portion from step (d) using a first and second sequencing primer;
- wherein the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer; a middle barcode portion comprising; and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides;
- wherein the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature;
- wherein the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer;
- wherein the first tail primer comprises a nucleic acid sequence identical to the tailed random primer; and
- wherein the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.
24. The method of paragraph 23, wherein the each tailed random primer further comprises a spacer nucleic acid sequence between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising about 6 to about 12 random nucleotides.
 25. The method of paragraph 23 or 24, wherein the unhybridized primers are removed from the reaction after an extension step.
 26. The method of any of paragraphs 23-25, wherein the second tail primer is nested with respect to the first tail primer by at least 3 nucleotides.
 27. The method of any of paragraphs 23-26, wherein the first target-specific primer further comprises a 5' tag sequence portion comprising a nucleic acid sequence of high GC content which is not substantially complementary to or substantially identical to any other portion of any of the primers.
 28. The method of any of paragraphs 23-27, wherein the second tail primer is identical to the full-length first sequencing primer.
 29. The method of any of paragraphs 23-28, wherein the portions of the target-specific primers that specifically anneal to the known target will anneal specifically at a temperature of about 65° C. in a PCR buffer.

50

30. The method of any of paragraphs 23-29, wherein the sample comprises genomic DNA.
31. The method of any of paragraphs 23-30, wherein the sample comprises RNA and the method further comprises a first step of subjecting the sample to a reverse transcriptase regimen.
32. The method of any of paragraphs 23-31, wherein the nucleic acids present in the sample have not been subjected to shearing or digestion or wherein the sample comprises single-stranded gDNA or cDNA.
33. The method of any of paragraphs 23-32, wherein the reverse transcriptase regimen comprises the use of random hexamers.
34. The method of any of paragraphs 23-33, wherein a gene rearrangement comprises the known target sequence.
35. The method of paragraph 34, wherein the gene rearrangement is present in a nucleic acid selected from the group consisting of: genomic DNA; RNA; and cDNA.
36. The method of any of paragraphs 34-35, wherein the gene rearrangement comprises an oncogene.
37. The method of paragraph 36, wherein the gene rearrangement comprises a fusion oncogene.
38. The method of any of paragraphs 23-37, wherein the nucleic acid product is sequenced by a next-generation sequencing method.
39. The method of paragraph 38, wherein the next-generation sequencing method comprises a method selected from the group consisting of: Ion Torrent, Illumina, SOLiD, 454; Massively Parallel Signature Sequencing solid-phase, reversible dye-terminator sequencing; and DNA nanoball sequencing.
40. The method of any of paragraphs 23-39, wherein the first and second sequencing primers are compatible with the selected next-generation sequencing method.
41. The method of any of paragraphs 23-40, wherein the method comprises contacting the sample, or separate portions of the sample, with a plurality of sets of first and second target-specific primers.
42. The method of any of paragraphs 23-41, wherein the method comprises contacting a single reaction mixture comprising the sample with a plurality of sets of first and second target-specific primers.
43. The method of any of paragraphs 23-42, wherein the plurality of sets of first and second target-specific primers specifically anneal to known target nucleotide sequences comprised by separate genes.
44. The method of any of paragraphs 23-43, wherein at least two sets of first and second target-specific primers specifically anneal to different portions of a known target nucleotide sequence.
45. The method of any of paragraphs 23-44, wherein at least two sets of first and second target-specific primers specifically anneal to different portions of a single gene comprising a known target nucleotide sequence.
46. The method of any of paragraphs 23-45, wherein at least two sets of first and second target-specific primers specifically anneal to different exons of a gene comprising a known nucleotide target sequence.
47. The method of any of paragraphs 23-46, wherein the plurality of first target-specific primers comprise identical 5' tag sequence portions.
48. The method of any of paragraphs 23-47, wherein each tailed random primer in a population of tailed random primers further comprises an identical sample barcoding portion.

US 10,450,597 B2

51

49. The method of paragraph 48, wherein multiple samples are each contacted with a separate population of tailed random primers with a sample barcoding portion; wherein each population of tailed random primers has a distinct sample barcoding portion; and wherein the samples are pooled after step (b).
50. The method of any of paragraphs 23-49, wherein each amplification step comprises a set of cycles of a PCR amplification regimen from 5 cycles to 20 cycles in length.
51. The method of any of paragraphs 23-50, wherein the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of from about 61 to 72° C.
52. The method of any of paragraphs 23-51, wherein the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of about 65° C.
53. The method of any of paragraphs 23-52, wherein the target nucleic acid molecule is from a sample, optionally which is a biological sample obtained from a subject.
54. The method of paragraph 53, wherein the sample is obtained from a subject in need of treatment for a disease associated with a genetic alteration.
55. The method of paragraph 54, wherein the disease is cancer.
56. The method of paragraph 53, wherein the sample comprises a population of tumor cells.
57. The method of paragraph 53, wherein the sample is a tumor biopsy.
58. The method paragraph 55, wherein the cancer is lung cancer.
59. The method of any of paragraphs 23-58, wherein a disease-associated gene comprises the known target sequence.
60. The method of 53, wherein a gene rearrangement product in the sample comprises the known target sequence.
61. The method of paragraph 60, wherein the gene rearrangement product is an oncogene.

EXAMPLES

Example 1: A Method Using Reverse Transcriptase with Tailed Random Oligonucleotides and Gene-Specific Oligonucleotides to Amplify 3' Fusion Events

First Strand Synthesis

As a first step towards amplifying 3' fusion events for sequence analysis, RNA was obtained from a sample isolated from a subject. The following reaction was assembled on ice to synthesize the first cDNA strand:

- 12 μ L purified RNA and H₂O
- 2 μ L 1.2 μ g/ μ L random primer #1 (9mer)
- 2 μ L dNTP

The reaction was transferred to a thermocycler and incubated at 65° C. for 5 minutes. Then, the reaction was centrifuged and incubated on ice for at least one minute.

- To the reaction above, the following was prepared on ice:
- 2 μ L 10 $\times$ M-MuLV reverse transcriptase buffer
 - 1 μ L 40 U/ μ L RNase inhibitor
 - 1 μ L 200 U/ μ L M-MuLV enzyme

52

The reaction was mixed and centrifuged briefly to collect the reaction contents at the bottom of the tube, then placed back on ice. The reaction was then incubated at 42° C. for 60 minutes, followed by 4° C.

ExoI Treatment

The following was added to the reaction above,

- 1 μ L 20 U/ μ L Exonuclease I

The reaction was mixed and centrifuged briefly to collect the contents at the bottom of the tube, then incubated at 37° C. for 10 minutes. Next, 1.28 μ L of 1N NaOH was added and mixed by pipetting up and down then centrifuged to collect the contents. The reaction was incubated at 80° C. for 10 minutes, then 4 μ L 10 mM Tris pH 8.3 was added and mixed by pipetting up and down. 20 μ L of the Exonuclease-treated DNA solution was transferred to a fresh 200 μ L PCR tube on ice.

Second Strand cDNA Synthesis

The following reaction was prepared:

- 20 μ L DNA solution, from above
- 11 μ L nuclease-free H₂O
- 4 μ L 10 $\times$ PCR Buffer II
- 4 μ L 3 μ M gene-specific primer #1
- 1 μ L 0.5 mM dNTP

The reaction was mixed by pipetting up and down then centrifuged briefly to collect the contents and placed on ice. The reaction was then incubated at 95° C. for 3 minutes, then 22° C. for 10 seconds followed by 4° C. until proceeding to the next step. The reaction was incubated on ice for at least one minutes.

The following was added to the reaction:

- 1 μ L 400 U/ μ L Manta 1.0 DNA Polymerase (high concentration)

The reaction was incubated at 25° C. for 10 seconds, then 70° C. for 10 minutes and maintained at 4° C. until proceeding to the next step.

DNA Purification with AMPure Beads #1

The following was added to the reaction above,

- 88.4 μ L AMPure beads

The suspension was mixed well and incubated for 5 minutes at room temperature. A magnet was used 2-4 minutes to collect the beads and the solution appeared clear. The supernatant was discarded and the beads were washed twice times with 200 μ L 70% ethanol on the magnet. After the second wash, the beads were dried at room temperature for 5 minutes. Finally, the DNA was eluted by removing the tubes from the magnet and resuspending the beads in 12 μ L 10 mM Tris-HCl pH 8.3 elution buffer included in the AMPure kit. The RNA-bead solution was placed on the magnet for 2 minutes. Then, the DNA solution was transferred to a fresh PCR tube, being sure to avoid transferring beads to the fresh tube.

It should be appreciated that in some embodiments the ratio of beads to reaction mix can effect the size of fragments returned. In some embodiments, for fusion detection, all or substantially all fragments are longer than 60 nt (e.g., 30 nt on either side of a fusion break point or junction) so each gene can be easily identified.

Amplification #1

The following reaction was prepared:

- 10 μ L purified DNA, from Purification #1 above
- 4 μ L 5 $\times$ Phoenix Hot Start Buffer
- 2 μ L 2 mM dNTP
- 2 μ L 10 μ M P5_barcode primer
- 2 μ L 3 μ M gene-specific primer #1
- 0.5 to 2 Units polymerase (e.g., Phoenix Hot Start Taq, VeraSeq)

US 10,450,597 B2

53

The reaction was incubated as follows:

Step 1: 95° C. for 3 minutes

Step 2: 95° C. for 30 seconds

Step 3: 65° C. for 5 minutes, return to step 2 for 14 cycles

Step 4: 72° C. for 2 minutes

Step 5: 4° C., until proceeding with the protocol

DNA Purification with AMPure Beads #2

The following was added to the reaction above,

36.4 µL AMPure beads

The suspension was mixed well and incubated for 5 minutes at room temperature. A magnet was used 2-4 minutes to collect the beads and the solution appeared clear. The supernatant was discarded and the beads were washed twice times with 200 µL 70% ethanol on the magnet. After the second wash, the beads were dried at room temperature for 5 minutes. Finally, the DNA was eluted by removing the tubes from the magnet and resuspending the beads in 9 µL 10 mM Tris-HCl pH 8.3 elution buffer included in the AMPure kit. The RNA-bead solution was placed on the magnet for 2 minutes. Then, the DNA solution was transferred to a fresh PCR tube, being sure to avoid transferring beads to the fresh tube.

Amplification #2

The following reaction was prepared:

8.5 µL purified DNA, from Purification #2 above

4 µL 5×Phoenix Hot Start Buffer

2 µL 2 mM dNTP

2 µL 10 µM P5_29 bp primer

2 µL 10 µM P7 barcode primer

2 µL 3 µM gene-specific primer #2

0.2 µL 5 U/µL Phoenix Hot Start Taq polymerase

2 µL 10 µM P7 barcode primer

The reaction was incubated as follows:

Step 1: 95° C. for 3 minutes

Step 2: 95° C. for 30 seconds

Step 3: 65° C. for 5 minutes, return to step 2 for 14 cycles

Step 4: 72° C. for 2 minutes

Step 5: 4° C., until proceeding with the protocol

DNA Purification with AMPure Beads #3

The following was added to the reaction above,

37.3 µL AMPure beads

The suspension was mixed well and incubated for 5 minutes at room temperature. A magnet was used 2-4 minutes to collect the beads and the solution appeared clear. The supernatant was discarded and the beads were washed twice times with 200 µL 70% ethanol on the magnet. After the second wash, the beads were dried at room temperature for 5 minutes. Finally, the DNA was eluted by removing the tubes from the magnet and resuspending the beads in 20 µL 10 mM Tris-HCl pH 8.3 elution buffer included in the AMPure kit. The RNA-bead solution was placed on the magnet for 2 minutes. Then, the DNA solution was transferred to a fresh PCR tube, being sure to avoid transferring beads to the fresh tube.

Quantification of Library Concentration

The Kapa Biosystems qPCR kit for Illumina was used to quantitate the concentration of each library prepared using the protocol above. The barcoded libraries were pooled at equimolar concentrations. Then the library was loaded on an Illumina MiSeq at XpM using the MiSeq v2 300 cycle reagent kit following the manufacturer's instruction. The samples were sequenced using 2×150 bp reads with 7 base encoded index reads.

Example 2: A Method Using Reverse Transcriptase with Gene-Specific Oligonucleotides and Tailed Random Oligonucleotides to Amplify 5' Fusion Events

First Strand Synthesis

As a first step towards amplifying 5' fusion events for sequence analysis, RNA was obtained from a sample iso-

54

lated from a subject. The following reaction was assembled on ice to synthesize the first cDNA strand:

12 µL purified RNA and H₂O

2 µL gene-specific primer #1

2 µL dNTP

The reaction was transferred to a thermocycler and incubated at 65° C. for 5 minutes. Then, the reaction was centrifuged and incubated on ice for at least one minute.

To the reaction above, the following was prepared on ice:

2 µL 10×M-MuLV reverse transcriptase buffer

1 µL 40 U/µL RNase inhibitor

1 µL 200 U/µL M-MuLV enzyme

The reaction was mixed and centrifuged briefly to collect the reaction contents at the bottom of the tube, then placed back on ice. The reaction was then incubated at 42° C. for 60 minutes, followed by 4° C.

Exol Treatment

The following was added to the reaction above,

1 µL 20 U/µL Exonuclease I

The reaction was mixed and centrifuged briefly to collect the contents at the bottom of the tube, then incubated at 37° C. for 10 minutes. Next, 1.28 µL of 1N NaOH was added and mixed by pipetting up and down then centrifuged to collect the contents. The reaction was incubated at 80° C. for 10 minutes, then 4 µL 10 mM Tris pH 8.3 was added and mixed by pipetting up and down. 20 µL of the Exonuclease-treated DNA solution was transferred to a fresh 200 µL PCR tube on ice.

Second Strand cDNA Synthesis

The following reaction was prepared:

20 µL DNA solution, from above

11 µL nuclease-free H₂O

4 µL 10×PCR Buffer II

4 µL 1.2 µg/µL random primer (9mer)

1 µL 0.5 mM dNTP

The reaction was mixed by pipetting up and down then centrifuged briefly to collect the contents and placed on ice. The reaction was then incubated at 95° C. for 3 minutes, then 22° C. for 10 seconds followed by 4° C. until proceeding to the next step. The reaction was incubated on ice for at least one minutes.

The following was added to the reaction:

1 µL 400 U/µL Manta 1.0 DNA Polymerase (high concentration)

The reaction was incubated at 25° C. for 10 seconds, then 70° C. for 10 minutes and maintained at 4° C. until proceeding to the next step.

DNA Purification with AMPure Beads #1

The following was added to the reaction above,

88.4 µL AMPure beads

The suspension was mixed well and incubated for 5 minutes at room temperature. A magnet was used 2-4 minutes to collect the beads and the solution appeared clear. The supernatant was discarded and the beads were washed twice times with 200 µL 70% ethanol on the magnet. After the second wash, the beads were dried at room temperature for 5 minutes. Finally, the DNA was eluted by removing the tubes from the magnet and resuspending the beads in 12 µL 10 mM Tris-HCl pH 8.3 elution buffer included in the AMPure kit. The RNA-bead solution was placed on the magnet for 2 minutes. Then, the DNA solution was transferred to a fresh PCR tube, being sure to avoid transferring beads to the fresh tube.

US 10,450,597 B2

55

Amplification #1

The following reaction was prepared:

10 μ L purified DNA, from Purification #1 above4 μ L 5 $\times$ Phoenix Hot Start Buffer2 μ L 2 mM dNTP2 μ L 10 μ M P5_barcode primer2 μ L 3 μ M gene-specific primer #1

0.5 to 2 Units polymerase (e.g., Phoenix Hot Start Taq, VeraSeq)

The reaction was incubated as follows:

Step 1: 95° C. for 3 minutes

Step 2: 95° C. for 30 seconds

Step 3: 65° C. for 5 minutes, return to step 2 for 14 cycles

Step 4: 72° C. for 2 minutes

Step 5: 4° C., until proceeding with the protocol DNA Purification with AMPure beads #2.

The following was added to the reaction above,

36.4 μ L AMPure beads

The suspension was mixed well and incubated for 5 minutes at room temperature. A magnet was used 2-4 minutes to collect the beads and the solution appeared clear. The supernatant was discarded and the beads were washed twice times with 200 μ L 70% ethanol on the magnet. After the second wash, the beads were dried at room temperature for 5 minutes. Finally, the DNA was eluted by removing the tubes from the magnet and resuspending the beads in 9 μ L 10 mM Tris-HCl pH 8.3 elution buffer included in the AMPure kit. The RNA-bead solution was placed on the magnet for 2 minutes. Then, the DNA solution was transferred to a fresh PCR tube, being sure to avoid transferring beads to the fresh tube.

Amplification #2

The following reaction was prepared:

8.5 μ L purified DNA, from Purification #2 above4 μ L 5 $\times$ Phoenix Hot Start Buffer2 μ L 2 mM dNTP

56

2 μ L 10 μ M P5_29 bp primer2 μ L 10 μ M P7 barcode primer2 μ L 3 μ M gene-specific primer #20.2 μ L 5 U/ μ L Phoenix Hot Start Taq polymerase2 μ L 10 μ M P7 barcode primer

The reaction was incubated as follows:

Step 1: 95° C. for 3 minutes

Step 2: 95° C. for 30 seconds

Step 3: 65° C. for 5 minutes, return to step 2 for 14 cycles

Step 4: 72° C. for 2 minutes

Step 5: 4° C., until proceeding with the protocol

DNA Purification with AMPure Beads #3

The following was added to the reaction above,

37.3 μ L AMPure beads

The suspension was mixed well and incubated for 5 minutes at room temperature. A magnet was used 2-4 minutes to collect the beads and the solution appeared clear. The supernatant was discarded and the beads were washed twice times with 200 μ L 70% ethanol on the magnet. After the second wash, the beads were dried at room temperature for 5 minutes. Finally, the DNA was eluted by removing the tubes from the magnet and resuspending the beads in 20 μ L 10 mM Tris-HCl pH 8.3 elution buffer included in the AMPure kit. The RNA-bead solution was placed on the magnet for 2 minutes. Then, the DNA solution was transferred to a fresh PCR tube, being sure to avoid transferring beads to the fresh tube.

Quantification of Library Concentration

The Kapa Biosystems qPCR kit for Illumina was used to quantitate the concentration of each library prepared using the protocol above. The barcoded libraries were pooled at equimolar concentrations. Then the library was loaded on an Illumina MiSeq at XpM using the MiSeq v2 300 cycle reagent kit following the manufacturer's instruction. The samples were sequenced using 2 $\times$ 150 bp reads with 7 base encoded index reads.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 5

<210> SEQ ID NO 1

<211> LENGTH: 7368

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

```

caagctttca agcattcaaa ggtctaaatg aaaaaggcta agtattattt caaaaggcaa      60
gtatatccta atatagcaaa acaaacaaaag caaatccat cagctactcc tocaattgaa     120
gtgatgaagc ccaaataatt catatagcaa aatggagaaa attagaccgg ccattctaaaa     180
atctgccatt ggtgaagtga tgaagaacat ttaactgttt attccgaagc ttgtcaattt     240
tgcaactctt ggctgccctat ggatttttgt ggtgcagtgt acagttttta atagetgect     300
aaagtcgtgt gtaactaatc tgggccagca gcttgacctt ggcaaccac ataacttgag     360
tgaaccgtgt atccaaggat gtcacttttg gaactctgta gatcagaaaa actgtgcttt     420
aaagtgtcgg gagtcgtgtg aggttggctg tagcagcgcg gaaggtgcat atgaagagga     480
agtactggaa aatgagagac taccaactgc tccctttgct tcttccattg gaagccacaa     540
tatgacatta cgatggaaat ctgcacactt ctctggagta aaatacatca ttcagtggaa     600
atatgcaaaa cttctgggaa gctggactta tactaagact gtgtccagac cgtctcatgt     660
ggtcaagccc ctgcacccct tcactgagta cattttccga gtggtttgga tcttcacagc     720

```


US 10,450,597 B2

57

58

- continued

gcagctgcag ctctactccc ctccaagtc cagttacagg actcactctc atggagttcc	780
tgaactgca cctttgatta ggaatattga gagctcaagt cccgacactg tggagtcag	840
ctgggaccca cctcaattcc caggtggacc tattttgggt tataacttaa ggctgatcag	900
caaaaatcaa aaattagatg cagggacaca gagaaccagt ttccagtttt actccacttt	960
accaaatact atctacaggt ttctattgc agcagtaaat gaagttgggt aggggtccaga	1020
agcagaatct agtattacca ctctactctc agcagttcaa caagaggaac agtggctctt	1080
tttatccaga aaaacttctc taagaaagag atctttaaaa catttagtag atgaagcaca	1140
ttgctctgg ttggatgcta tataccataa tattacagga atatctgttg atgtccacca	1200
gcaaatgtt tatttctctg aaggaactct catatgggct aagaaggtcg ccaacatgtc	1260
tgatgtatct gacctgagaa ttttttacag aggttcagga ttaatttctt ctatctccat	1320
agattggctt tatcaaaaga tgtatttcat catggatgaa ctggatatgtg tetgtgattt	1380
agagaactgc tcaaacatcg aggaattac tccacctctt attagtgcac ctcaaaaaat	1440
tgtggctgat tcatacaatg ggtatgtctt ttacctctg agagatggca ttatagagc	1500
agaccttctt gtaccatctg gccggtgtgc agaagctgtg cgtattgtgg agagttgcac	1560
gttaaaggac ttgtcaatca agccacaagc caagcgaatc atttacttca atgacactgc	1620
ccaagtcttc atgtcaacat ttctggatgg ctctgcttcc catctcatcc tactctgcat	1680
cccttttgc tggatgaaaa gttttgcttg tgaatacaat gaactttctg tcacagatgg	1740
caaggtcatc ttccaacagg atgctttgtc tttaaatgaa ttcacgtgg gatgtgacct	1800
gagtcacata gaagaatttg ggtttggtaa cttgggtcatc ttgggtcat cctccagct	1860
gcacctctg ccagggccgc cgcaggagct ttcggtgctg ttggctctc accaggtctt	1920
tgttcaatgg aagctctctg ccttgccat agggaccaat gtcacctga tcagtgatat	1980
tattgaactc ttgaattag gcccttctgc ctggcagaa cggacctatg aggtgaaagt	2040
atccacccaa gacctctg aagtcactca tattttcttg aacataagtg gaaccatgct	2100
gaatgtacct gagctgcaga gtgctatgaa atacaagggt tctgtgagag caagtctctc	2160
aaagaggcca ggcctctgt cagagccctc agtgggtact accctgggtc cagctagtga	2220
accaccattt atcatggctg tgaagaaga tgggcttttg agtaaacat taaatagctt	2280
tggccagga gagttcttat cctctgatat aggaaatgtg tcagacatgg attgggtataa	2340
caacagctct tactacagt acacgaaagg cgaagttttt gtgtggtctg tgaatgggac	2400
ggatatctca gagaattatc acctaccag cattgcagga gcaggggctt tagcttttga	2460
gtggctgggt cactttctct actgggctgg aaagacatat gtgatacaaa ggcagttctg	2520
gttgacggga cacacagaca ttgttacc cagtgaagcta ttggatgaatg acatgggtgt	2580
ggattcagtt ggtggatctc tctactggac cactctctat tcagtgga aa gccacagact	2640
aaatggggaa agttcccttg tactacagac acagccttgg tttctctgga aaaaggtaat	2700
tgtcttaact ttagacctca gtgatgggt cctgtatttg ttggttcaag acagtcaatg	2760
tattcacctg tacacagctg ttcttgggg acagagcact ggggatacca ccatcacaga	2820
atttgcagcc tggagtactt ctgaaatttc ccagaatgca ctgatgtact atagtggctg	2880
gtgtttcttg atcaatggct ttaggattat cacaactcaa gaaataggct agaaaaccag	2940
tgtctctgtt ttggaaccag ccagatttaa tcagttcaca attattcaga catcccttaa	3000
gccccgcca gggaaacttt cctttacccc taagggtatt ccagattctg ttcaagagtc	3060

US 10,450,597 B2

59

60

- continued

ttcatttagg attgaaggaa atgcttcaag ttttcaaat	ctgtggaatg gteccctgc	3120
ggtagactgg ggtgtagtgt tctacagtgt agaatttagt	gctcattcta agttcttggc	3180
tagtgaacaa cactctttac ctgtattttac tgtggaagga	ctggaacctt atgccttatt	3240
taatctttct gtcactcctt atacctactg gggaaagggc	cccaaaacat ctctgtcact	3300
tcgagcacct gaaacagttc catcagcacc agagaacecc	agaatattta tattaccaag	3360
tggaaaatgc tgcaacaaga atgaagttgt ggtggaattt	agggtggaaca aacctaaagca	3420
tgaaaatggg gtgttaacaa aatttgaaat tttctacaat	atatccaatc aaagtattac	3480
aaacaaaaca tgtgaagact ggattgctgt caatgtcact	ccctcagtga tgtcttttca	3540
acttgaaggc atgagtcacca gatgctttat tgcttccag	gttagggcct ttaçatctaa	3600
ggggccagga ccatatgctg acgttgtaaa gtctacaaca	tcagaaatca acccatttcc	3660
tcacctcata actcttcttg gtaacaagat agttttttta	gatatggatc aaaatcaagt	3720
tgtgtggacg ttttcagcag aaagagttat cagtgcctgt	tgctacacag ctgataatga	3780
gatgggatat tatgtgaag gggactcact ctctctctctg	cacttgcaca atcgtcttag	3840
ctctgagctt ttccaagatt cactgggttt tgatatcaca	gttattacaa ttgactggat	3900
ttcaaggcac ctctactttg cactgaaaga atcacaaat	ggaatgcaag tatttgatgt	3960
tgatcttgaa cacaaggtga aatatcccag agaggtgaag	attcacata ggaattcaac	4020
aataatttct ttttctgtat atctctttt aagtctcttg	tattggacag aagtttccaa	4080
ttttggctac cagatgttct actacagtat tatcagtcac	accttgcacc gaattctgca	4140
accacagct acaaaccaac aaacaaaaag gaatcaatgt	tcttgtaatg tgaactgaatt	4200
tgagttaagt ggagcaatgg ctattgatac ctctaacct	gagaaacct tgatatactt	4260
tgccaaagca caagagatct gggcaatgga tetggaagge	tgtcagtggt ggagagttat	4320
cacagtacct gctatgctcg caggaaaaac ccttggttagc	ttactgtgg atggagatct	4380
tatatactgg atcatcacag caaaggacag cacacagatt	tatcaggcaa agaaggaaa	4440
tggggccatc gtttcccagg tgaaggccct aaggagtagg	catatcttgg cttacagttc	4500
agttatgcag ccttttccag ataaagcgtt tetgtctcta	gcttcagaca ctgtggaacc	4560
aactatactt aatgccacta acactagcct cacaatcaga	ttacctctgg ccaagacaaa	4620
cctcacatgg tatggcatca ccagccctac tccaacatac	ctggtttatt atgcagaagt	4680
taatgacagg aaaaacagct ctgacttgaa atatagaatt	ctggaatttc aggacagtat	4740
agctcttatt gaagattttc aaccattttc aacatacatg	atacagatag ctgtaaaaaa	4800
ttattattca gatccttttg aacattttacc accaggaaa	gagatttggg gaaaaactaa	4860
aaatggagta ccagaggcag tgcagctcat taatacaact	gtgcggtcag acaccagcct	4920
cattatater tggagagaat ctcaacagcc aaatggacct	aaagaatcag tccgttatca	4980
gttggaatc tcacccctgg cctcaattcc tgaaaactcc	ctaagacaaa gtgaatttcc	5040
aaatggaagg ctcaactctc ttgttactag actgtctggt	ggaaatattt atgtgttaaa	5100
ggttcttgcc tgccactctg aggaaatgtg gtgtacagag	agtcactctg tcaactgtga	5160
aatgtttaac acaccagaga aaccttattc cttgggtcca	gagaacacta gtttgcaatt	5220
taattggaag gctccattga atgttaacct catcagattt	tgggttgagc tacagaagtg	5280
gaaatacaat gagttttacc atgttaaaac ttcattgcagc	caaggctctg cttatgtctg	5340
taatatcaca aatctacaac ctctacttc atataatgtc	agagtagtgg tggtttataa	5400
gacgggagaa aatagcacct cacttccaga aagctttaag	acaaaagctg gagtcccaaa	5460

US 10,450,597 B2

61

62

- continued

taaaccaggc attcccaaat tactagaagg gagtaaaaat tcaatacagt gggagaaaagc	5520
tgaagataat ggatgtagaa ttacatacta tatecttgag ataagaaaaga gcacttcaaa	5580
taattttacag aaccagaatt taagggtgaa gatgacattt aatggatcct gcagtagtgt	5640
ttgcacatgg aagtccaaaa acctgaaaagg aatatttcag ttcagagtag tagctgcaaa	5700
taatctaggg ttgtgtgaat atagtggaat cagtggagaat attatattag ttggagatga	5760
tttttggata ccagaaacaa gtttcatact tactattata gttggaatat ttctgttgt	5820
tacaatccca ctgacctttg tctggcatag aagattaaag aatcaaaaaa gtgccaagga	5880
aggggtgaca gtgcttataa acgaagacaa agagttggct gagctgcgag gtctggcagc	5940
cggagtaggc ctggctaagt cctgctatgc aatacatact ctcccaaccc aagaggagat	6000
tgaaaatctt cctgccttcc ctgggaaaa actgactctg cgtctcttgc tgggaagtgg	6060
agcctttgga gaagtgtatg aaggaaacagc agtggacatc ttaggagttg gaagtggaga	6120
aatcaaaagta gcagtgaaga ctttgaagaa gggttccaca gaccaggaga agattgaatt	6180
cctgaaggag gcacatctga tgagcaaat taatcatccc aacattctga agcagcttgg	6240
agtttgtctg ctgaatgaac cccaatacat tatcctggaa ctgatggagg gaggagacct	6300
tcttacttat ttgcgtaaag cccggatggc aacgttttat ggtcctttac tcaccttgg	6360
tgaccttgta gacctgtgtg tagatatctc aaaaggctgt gtctacttgg aacggatgca	6420
tttcaatcac agggatctgg cagctagaaa ttgccttgtt tccgtgaaa actataccag	6480
tcacggata gtgaagattg gagactttgg actcgccaga gacatctata aaaatgatta	6540
ctatagaaag agaggggaag gcctgctccc agtctgggtg atggctccag aaagtttgat	6600
ggatggaatc ttcactactc aatctgatgt atggtctttt ggaattctga ttggggagat	6660
tttaactctt ggtcatcage cttatccagc tcattccaac cttgatgtgt taaactatgt	6720
gcaaacagga gggagactgg agccaccaag aaattgtcct gatgatctgt ggaatttaat	6780
gacccagtgc tgggctcaag aacccgacca aagacctact ttccatagaa ttcaggacca	6840
acttcagtta ttcagaaaat ttttcttaa tagcatttat aagtcacag atgaagcaaa	6900
caacagtgga gtcataaatg aaagctttga aggtgaagat ggcgatgtga ttgtttgaa	6960
ttcagatgac attatgccag ttgctttaat ggaacgaag aaccgagaag ggttaacta	7020
tatggtactt gctacagaat gtggccaaag tgaagaaaag tctgagggtc ctctagggtc	7080
ccaggaaatc gaatcttgtg gtctgaggaa agaagagaag gaaccacatg cagacaaaag	7140
ttcttgccaa gaaaaacaag tggcttactg cccttctggc aagcctgaag gcttgaacta	7200
tgctgtctc actcacagtg gatatggaga tgggtctgat taatagcgtt gtttgggaaa	7260
tagagagttg agataaacac tctcattcag tagttactga aagaaaaetc tgctagaatg	7320
ataaatgtca tgggtgtcta taaactcaaa taaacaatgc aacgttcc	7380

<210> SEQ ID NO 2
 <211> LENGTH: 6267
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

agctgcaagt ggcgggcgcgc caggcagatg cgatccagcg gctctggggg cggcagcgg	60
ggtagcagct ggtacctccc gccgcctctg ttggagggtt cgcggggcac cgaggtgctt	120
tccggccgcgc ctctggctgg ccacccaaa cgcggggcgc tgatgatggg tggaggagg	180

US 10,450,597 B2

63

64

- continued

gcggaagat ttggggcgcc cctgcccga acgcccag ctgctgcgc cggggcgcgt	240
ccagtgcctg cgaactctga ggagccgagg cgccgggtgag agcaaggacg ctgcaaaactt	300
gagcagcgcg ggggctggga ttacgcccc gaagttcagc aggcagacag tccgaagcct	360
tcccgacgag gagagatagc ttgagggtgc gcaagacggc agcctccgc ctcggttccc	420
gccagaccg ggcagaagag cttggaggag ccaaaaaggaa cgcaaaagge ggccaggaca	480
gcgtgcagca gctgggagcc gccgttctca gccctaaaag ttgcagagat tggaggctgc	540
cccgagaggg gacagacccc agctccgact gcggggggca ggagaggacg gtacccaact	600
gccacccccc ttcaaccata gtagtctctc tgtaccgagc gcagcgagct acagacgggg	660
gcgcggcact cggcgcgagg agcgggagge tcaaggctcc agccagttag cccagtgtgc	720
ttgagtgtct ctggactcgc ccctgagctt ccaggctctgt ttcatttaga ctctgtctcg	780
cctccgtgca gttgggggaa agcaagaac ttgcgcgcac gcacagctct ctggagatca	840
ggtggaaggg gccgctgggt acaaggact gttcagagcc tcttcccatc tcggggagag	900
cgaagggtga ggcctgggcc ggagagcagt gtaaacggcc tctccggcg ggatgggagc	960
cctcgggctc ctgtggctcc tgcgctgct gctttccacg gcagctgtgg gctccgggat	1020
ggggacccgc cagcgcgagg gctcccagc tggggggccg ccgctgcagc cccgggagcc	1080
actcagctac tcgcgctcgc agagggaagag tctggcagtt gacttctgtg tgcctcgcct	1140
cttcctgtgc taaccccgcc acctactgct gccaccatcc tctcggagc tgaaggctgg	1200
caggcccgag gcccgcggtc cgtagctct ggactgcgc ccgctgctca ggttgctggg	1260
gccggcgccg ggggtctctt ggaccgccg ttccaccagc ccggcagagg cccggacgct	1320
gtccagggtg ctgaaggcg gctccgtgcg caagctccgg cgtgccaaag agttggtgct	1380
ggagctgggc gaggaggcga tcttgagggg ttgcgtcggg ccccccgggg aggcggctgt	1440
ggggctgctc cagtccaatc tcagcgagct gttcagttgg tggattcgcc aaggcgaagg	1500
gcgactgagg atccgctcta tgcccagaa gaaggcgtcg gaagtgggca gagagggaag	1560
gctgtccgcg gcaattcgcg cctcccagcc ccgcttctc ttccagatct tcgggactgg	1620
tcatagtccc ttggaatcac caacaaacat gcccttctct tctcctgatt attttacatg	1680
gaatctcacc tggataatga aagactcctt ccttttctg tctcctgca gccgatatgg	1740
tctggagtgc agctttgact tcccctgtga gctggagtat tcccctccac tgcattgact	1800
caggaaaccg agctggctct ggcgccgcat cccctccgag gaggcctccc agatggactt	1860
gctggatggg cctggggcag agcgttctaa ggagatgccc agaggctcct ttctcttct	1920
caacacctca gctgactcca agcacaccat cctgagtcgc tggatgagga gcagcagtga	1980
gcactgcaca ctggccgtct cggcgacag gcacctgcag cctctggaa ggtacattgc	2040
ccagctgctg ccccaaacg aggctgcaag agagatcctc ctgatgccca ctccaggga	2100
gcattggttg acagtgtccc agggaagaat cgggcgtcca gcaaacccat ttcgagtggc	2160
cctggaatac atctccagtg gaaaccgcag cttgtctgca gtggactttt ttgcctgaa	2220
gaactgcagt gaaggaaacat ccccaggctc caagatggcc ctgcagagct ccttcaactg	2280
ttggaatggg acagtctccc agcttgggca ggctgtgac ttccaccagg actgtgccca	2340
gggagaagat gagagccaga tgtgocggaa actgctgtg ggtttttact gcaactttga	2400
agatggcttc tgtggctgga cccaaggcac actgtcacc cactactcct aatggcaggt	2460
caggacccta aaggatgccc ggttccagga ccaaccaagc catgctctat tgctcagtac	2520
cactgatgtc cccgcttctg aaagtgtcac agtgaccagt gctacgttct ctgcaccgat	2580

US 10,450,597 B2

65

66

- continued

caagagctct ccatgtgagc tccgaatgtc ctggetcatt cgtggagtct tgaggggaaa	2640
cgtgtccttg gtgctagtgg agaacaaaac cgggaaggag caaggcagga tggctctggca	2700
tgtgcgcgc tatgaaggct tgagcctgtg gcagtggatg gtgttgcttc tctctgatgt	2760
gtctgacagg ttctggctgc agatggctgc atggctggga caaggatcca gagccatcgt	2820
ggcttttgac aatatctcca tcagcctgga ctgctacctc accattagcg gagaggacaa	2880
gacccctgcag aatacagcac ccaaatcaag aaacctgttt gagagaaaacc caaacaagga	2940
gctgaaaacc ggggaaaatt caccagaca gacccccate ttgacccta cagttcattg	3000
gctgttcacc acatgtgggg ccagcgggccc ccatggcccc acccaggcac agtgcaacaa	3060
cgcctaccag aactccaacc tgagcgtgga ggtggggagc gaggggcccc tgaaggcat	3120
ccagatctgg aagggtgccac ccaccgacac ctacagcacc tcgggctacg gactgctgg	3180
cgggaaaggc gggagaagaa ccatgatgcg gtccacggc gtgtctgtgc tgggcatctt	3240
caacctggag aaggatgaca tgctgtacat cctggttggg cagcaggag aggacgcctg	3300
ccccagtaca aaccagtaa tccagaaagt ctgcattgga gagaacaatg tgatagaaga	3360
agaaatccgt gtgaacagaa gcgtgcatga gtgggcagga ggcggaggag gagggggtgg	3420
agccacctac gtatttaaga tgaaggatgg agtgccgggtg cccctgatca ttgcagccgg	3480
aggtggtggc agggcctacg gggccaagac agacacgttc caccagaga gactggagaa	3540
taactcctcg gttctagggc taaacggcaa ttccggagcc gcaggtgggtg gaggtggctg	3600
gaatgataac acttctctgc tctgggcccgg aaaatctttg caggagggtg ccaccggagg	3660
acattctctg cccaggccca tgaagaagtg ggggtgggag acaagagggg gtttcggagg	3720
gggtggaggg ggggtgctct caggtggagg aggcggaggga tatataggcg gcaatgcagc	3780
ctcaaacaaat gaccccgaaa tggatgggga agatgggggt tcttctatca gtccactggg	3840
cactctgtac accccagctt taaaagtgat ggaaggccac ggggaagtga atattaagca	3900
ttatctaaac tgacgtcact gtgaggtaga cgaatgtcac atggacctg aaagccacaa	3960
ggtcatctgc ttctgtgacc acgggacggg gctggctgag gatggcgtct cctgcattgt	4020
gtcaccacac ccggagccac acctgccact ctcgctgac ctctctgtgg tgacctctgc	4080
cctcgtggcc gccctggctc tggctttctc cggcatcatg attgtgtacc gccggaagca	4140
ccaggagctg caagccatgc agatggagct gcagagccct gagtacaagc tgagcaagct	4200
ccgcacctcg accatcatga ccgactacaa ccccaactac tgetttgctg gcaagaacctc	4260
ctccatcagt gacctgaagg aggtgcctcg gaaaaacate accctaatc ggggtctggg	4320
ccatggcgcc ttctggggagg tgtatgaagg ccagggtgtc ggaatgccc acgacccaag	4380
ccccctgcaa gtggctgtga agacgctgcc tgaagtgtgc tctgaacagg acgaaactgga	4440
tttctctatg gaagccctga tcatcagcaa attcaaccac cagaacattg ttcgctgcat	4500
tgggtgagc ctgcaatccc tgcccgggtt cactctgctg gactcatgg cggggggaga	4560
cctcaagtcc ttcctccgag agaccggccc tcgcccagc cagccctcct cctggccat	4620
gctggacctt ctgcacgtgg ctgggacat tgctgtggc tgtcagtatt tggaggaaaa	4680
ccacttcac caccgagaca ttgctgccag aaactgcctc ttgacctgtc caggccctgg	4740
aagagtggc aagattggag acttcgggat ggcgcgagac atctacaggg cgaactacta	4800
tagaaagga ggcctgtcca tgctgccagt taagtggatg cccccagagg ccttcatgga	4860
aggaatatcc acttctaaaa cagacacatg gtctcttggg gtgctgctat gggaaatctt	4920

US 10,450,597 B2

67

68

- continued

```

ttctcttggga tatatgccat accccagcaa aagcaaccag gaagttcttg agtttgtcac 4980
cagtggaggc cggatggacc cacccaagaa ctgccctggg cctgtatacc ggataatgac 5040
tcagtgtctg caacatcage ctgaagacag gcccactttt gccatcattt tggagaggat 5100
tgaatactgc acccaggacc cggatgtaat caacaccgct ttgccgatag aatatggtec 5160
acttgtggaa gaggaagaga aagtgcctgt gaggccdaag gacctgagg gggttctctc 5220
tctcctggtc tctcaacagg caaaacggga ggaggagcgc agcccagctg ccccaaccacc 5280
tctgcctacc acctcctctg gcaaggctgc aaagaaaccc acagctgcag agatctctgt 5340
tcgagtcctt agagggccgg ccgtggaagg gggacacgtg aatatggcat tctctcagtc 5400
caaccctctt tcggagtgc acaaggtcca cggatccaga aacaagccca ccagcttgtg 5460
gaacccaacg tacggctcct ggtttacaga gaaacccacc aaaaagaata atcctatagc 5520
aaagaaggag ccacacgaca ggggtaacct ggggctggag ggaagctgta ctgtccacc 5580
taacgttgca actgggagac ttccgggggc ctcaactgtc cttagagcctt cttegtgac 5640
tgccaatatg aaggaggta cctctgttcag gctacgtcac ttcccttggt ggaatgtcaa 5700
ttacggctac cagcaacagg gcttgcctt agaagccgt actgcccctg gagctggtea 5760
ttacgaggat accattctga aaagcaagaa tagcatgaac cagcctgggc cctgagctcg 5820
gtcgcacact cacttctctt ccttgggato cctaagaccg tggaggagag agaggcaatg 5880
gctccttcac aaaccagaga ccaaatgtca cgttttgtt tgtgccaaac tattttgaag 5940
taccacaaaa aaagctgtat ttgaaaaatg ctttagaaag gttttgagca tgggttcac 6000
ctattctttc gaaagaagaa aatatcataa aaatgagtga taaatacaag gccagatgt 6060
gggtgcataa ggtttttatg catgtttgt gtatacttc ttatgcttct ttc aaattgt 6120
gtgtgctctg etteaagtga gtcagaatta gctgcttcta tgttteatag ttggggtcac 6180
agatgttttc ttgccttgt gatgtggaca tgagccattt gaggggagag ggaacggaaa 6240
taaaggagtt atttgaatg actaaaa 6267

```

```

<210> SEQ ID NO 3
<211> LENGTH: 4174
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 3

```

```

agtcgccga ccgaagcagg gcgcgcagca gcgtgagtg ccccggaacg tgcgtcgcgc 60
cccagtgtc cgtcgcgtcc gccgcgcgcc gggcggggat gggcgggcca gactgagcgc 120
cgcacccgcc atccagaccc gccggcccta gccgcagtc ctcagccgt ggcaccagcg 180
cgcacgggag atggcgaagg cgacgtccgg tgcgcgggg ctgcgtctgc tgttgcgtct 240
gctgctgcc ctgctaggca aagtggcatt gggcctctac ttctcgagg atgcttactg 300
ggagaagctg tatgtggacc agggggccgg caccgccctg ctgtacgtcc atgcccgtgc 360
ggacgcccct gaggagggtc ccagcttccg cctgggccag catctctacg gcaagtaacc 420
cacacggctg catgagaaca actggatctg catccaggag gacacgggcc tctctacct 480
taaccggagc ctggaccata gctcctggga gaagctcagt gtccgcaacc gcggctttcc 540
cctgctcacc gtctacctca aggtcttctt gtcacccaca tcccttcgtg agggcgagtg 600
ccagtgggca ggctgtgccc gcgtataact ctcttcttcc aacacctcct ttcagacctg 660
cagctccctc aagccccggg agctctgctt cccagagaca aggcctcct tccgcatctg 720
ggagaaccga cccccagga ccttcacca gttccgctg ctgctgtgc agttcttgtg 780

```


US 10,450,597 B2

69

70

- continued

ccccaacatc	agcgtggcct	acaggtctct	ggaggggtgag	ggtctgcctt	tccgtgcgc	840
cccgacacgc	ctggaggtga	gcacgcgctg	ggcctggac	cgcgagcagc	gggagaagta	900
cgaactggtg	gcccgtgtca	cgcgtcacgc	cggcgcgcg	gaggaggtgg	tgatggtgcc	960
cttcccgggtg	acccgtgtacg	acgaggacga	ctcgccgcct	accttccccg	cgggctgca	1020
caccgccacg	gcccgtgtgg	agttcaagcg	gaaggaggac	accgtggtgg	ccacgctgcg	1080
tgtcttcgat	gcagacgtgg	tacctgcac	aggggagctg	gtgagggcgt	acacaagcac	1140
gctgctcccc	ggggacacct	gggcccagca	gaccttcggg	gtggaacct	ggcccaacga	1200
gacctcggtc	caggccaacg	gcagcttcgt	gcggggcgacc	gtacatgact	ataggtggt	1260
tctcaaccgg	aacctctcca	tctcgagaa	cgcacccatg	cagctggcgg	tgtggtcaa	1320
tgactcagac	tccagggccc	caggagcggg	cgtcctcttg	ctccacttca	acgtgtcgg	1380
gctgcccgtc	agcctgcacc	tgcctcagta	ctactccttc	tccgtgagca	ggagggctcg	1440
ccgatttgcc	cagatcgagg	aagtctgtgt	ggaaaactgc	caggcattca	gtggcatcaa	1500
cgtccagta	aagctgcatt	cctctgggtg	caactgcagc	acgctagggg	tggctacctc	1560
agccgaggac	acctcgggga	tctgtttgt	gaatgacacc	aaggccctgc	ggcgcccaa	1620
gtgtgccgaa	cttcactaca	tgggtgtggc	caccgaccag	cagacctcta	ggcaggccca	1680
ggcccagctg	cttgtaacag	tggaggggtc	atatgtggcc	gaggaggcgg	gctgccccct	1740
gtcctgtgca	gtcagcaaga	gacggctgga	gtgtgaggag	tgtggcgccc	tgggtctccc	1800
aacaggcagg	tgtgagtgga	ggcaaggaga	tggcaaaggg	atcaccagga	acttctccac	1860
ctgctctccc	agcaccaga	cctgccccga	cggccactgc	gatgttgtgg	agaccaaga	1920
catcaacatt	tgcctcagg	actgcctcgg	gggcagcatt	gttgggggac	acgagcctgg	1980
ggagcccccg	gggattaaag	ctggctatgg	cacctgcaac	tgcttccctg	aggaggagaa	2040
gtgcttctgc	gagcccgaa	acatccagga	tccactgtgc	gacgagctgt	gccgcacgg	2100
gatcgcagcc	gctgtctctc	tctcttcat	cgtctcgggt	ctgctgtctg	ccttctgcac	2160
ccactgtac	cacaagtttg	ccacaagcc	acccatctcc	tcagctgaga	tgaccttccg	2220
gaggccccgc	caggcccttc	cggctagcta	ctctcttcc	ggtgcgcgc	ggccctcgct	2280
ggactccatg	gagaaccagg	tctcctgtga	tgccttcaag	atcctggagg	atccaaagt	2340
ggaattccct	cggagaagct	tggttcttgg	aaaaactcta	ggagaaggcg	aatttgga	2400
agtggtaag	gcaacggcct	tccatctgaa	aggcagagca	gggtacacca	cgggtggcgt	2460
gaagatgctg	aaagagaacg	cctccccgag	tgagcttcca	gacctgctgt	cagagttcaa	2520
cgtcctgaag	caggtaaccc	acccacatgt	catcaaatgt	tatggggcct	gcagccagga	2580
tggcccgctc	ctcctcatcg	tggagtacgc	caaatacggc	tccctggcgg	gcttctctcg	2640
cgaagacccg	aaagtggggc	ctggctacct	gggcagtgga	ggcagccgca	actccagctc	2700
cctggaccac	cggatgagc	gggcctcac	catgggcgac	ctcatctcat	ttgcctggca	2760
gatctccag	gggatgcagt	atctggccga	gatgaagctc	gttcacggcg	acttggcagc	2820
cagaaacatc	ctggtagctg	agggggcgaa	gatgaagatt	tggatttctg	gcttgcctcg	2880
agatgtttat	gaagaggatt	cctacgtgaa	gaggagccag	ggtcggtatc	cagttaaagt	2940
gatggcaatt	gaatcccttt	ttgatcatat	ctacacccacg	caaagtgatg	tatggctctt	3000
tgggtctctg	ctgtgggaga	tctgtaccct	agggggaaac	ccctatctcg	ggattctctc	3060
tgagcggctc	ttcaaccttc	tgaagaccgg	ccaccggatg	gagagggcag	acaactgcag	3120

US 10,450,597 B2

71

72

- continued

```

cgaggagatg taecgectga tgctgcaatg ctggaagcag gagccggaca aaaggccggg 3180
gtttgccggac atcagcaaaag acctggagaa gatgatgggt aagaggagag actacttgga 3240
cettgcccgg tccactccat ctgactccct gatttatgac gacggccctct cagaggagga 3300
gacaccgctg gtggactgta ataatgcccc cctccctega gccctccctt ccacatggat 3360
tgaaaacaaa ctctatggta gaatttccca tgcatttact agattctage accgctgtcc 3420
cctctgcaat atccttccct tctgtgatgc ttttataaaa tgtttctggg ctgaacaaaa 3480
ccaaagtctg ctctgaacct ttttatttgt aaatgtctga ctttgcaccc agtttacatt 3540
taggcattat tgcaactatg tttttctaaa aggaagtga aataagtga attaccacat 3600
tgcccagcaa cttaggatgg tagaggaaaa aacagatcag ggcggaaact tcaggggaga 3660
ccaagaacag gttgaataag gcgcttctgg ggtgggaatc aagtcatagt acttctactt 3720
taactaagtg gataaatata caaatctggg gaggtattca gttgagaag gagccaccag 3780
caccactcag cctgcaactg gagcacagcc aggttccccc agacccctcc tgggcaggca 3840
ggtgcctctc agagggccac cgccactggc gagcagccac tggccaagcc tcagcccag 3900
tcccagccac atgtccctca tcaggggtag cgaggttgca ggagctggct ggccctggga 3960
ggacgcacc ccaactgtgt tttcacatcc tttcccttac ccaccttcag gacgggtgtc 4020
acttatgaag tcagtgtctaa agctggagca gttgcttttt gaaaagaacat ggtctgtggt 4080
gctgtggtct tacaatggac agtāaatatg gttcttgcca aaactccttc ttttgtcttt 4140
gattaaatac tagāaattta aaaaaaaaaa aaaa 4174

```

<210> SEQ ID NO: 4

<211> LENGTH: 5629

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

```

agtcccgca ccgaagcagg gcgcgcagca gcgctgagtg ccccggaacg tgcgtcgcgc 60
cccagtgtc cgtcgcgtcc gccgcgcccc gggcggggat gggcgggcca gactgagcgc 120
cgcacccgcc atccagaccc gccggcccta gccgcagtcc ctccagccgt ggcccagcgc 180
cgcacggggc atggcgaagg cgacgtccgg tgcgcggggg ctgctgttgc tgttctgtct 240
gctgtctccg ctgctaggca aagtgccatt gggcctctac ttctcgaggg atgcttactg 300
ggagaagctg tatgtggacc aggcggccgg caccgccctg ctgtacgtcc atgccctgcg 360
ggacgcccc ctgagggtgc ccagcttccg cctgggccag catctctacg gcaagtaacc 420
cacacggctg catgagaaca actggatctg catccaggag gacacgggcc tctctacact 480
taaccggagc ctggaccata gctcctggga gaagctcagt gtcgcgaacc gcggctttcc 540
cctgtccacc gtctacctca aggtcttctc gtcacccaca tccctctgtg agggcgagtg 600
ccagtggcca ggctgtgccc gcttatactt ctccctcttc aacacctctt ttccagcctg 660
cagctccctc aagcccccgg agctctgctt cccagagaca aggcctctct tccgcattcg 720
ggagaaccga ccccaggca ccttcacca gtccgcctg ctgctgtgc agttcttctg 780
ccccaacatc agcgtggcct acaggctcct ggagggtgag ggtctgcctt tccgtgcgc 840
cccggacagc ctggagggtg gcacgcgctg ggccttgac cgcgagcagc gggagaagta 900
cgagctggtg gccgtgtgca ccgtgcacgc cggcgcgccg gaggaggtgg tgatggtgcc 960
cttcccgtg accgtgtacg acgaggacga ctccggcccc accttccccg cgggcgtcga 1020
cacgcacagc gccgtggtgg agttcaagcg gaaggaggac accgtggtgg ccacgtgcgc 1080

```


US 10,450,597 B2

73

74

- continued

tgtcttcgat	gcagacgtgg	tacctgcac	aggggagctg	gtgaggcgg	acacaagcac	1140
gctgctcccc	ggggacacct	ggcccagca	gaccttcgg	gtggaacct	ggcccaacga	1200
gacctcggtc	caggccaacg	gcagcttcgt	gcgggcgacc	gtacatgact	ataggctggt	1260
tetcaaccgg	aacctctcca	tctcgagaa	cgcaccatg	cagctggcgg	tgtggtcaa	1320
tgactcagac	ttccagggcc	caggagcggg	cgtectcttg	ctccacttca	acgtgtcgg	1380
gctgccggtc	agcctgcacc	tgcacgtac	ctactccctc	tccgtgagca	ggagggtcgt	1440
cggatttgcc	cagatcgga	aagtctgtgt	ggaaaactgc	caggcattca	gtggcatcaa	1500
cgtccagtac	aagctgcatt	cctctggtgc	caactgcage	acgttagggg	tggtcacctc	1560
agccgaggac	acctcgggga	tcctgtttgt	gaatgacacc	aaggccctgc	ggcgcccaa	1620
gtgtgccgaa	cttcactaca	tgggtgtggc	caccgaccag	cagacctcta	ggcaggccca	1680
ggcccagctg	cttgtaacag	tggaggggtc	atatgtggcc	gaggaggcgg	gctgccccct	1740
gtcctgtgca	gtcagcaaga	gacggctgga	gtgtgaggag	tgtggcggcc	tgggtcccc	1800
aacaggcagg	tgtgagtga	ggcaaggaga	tggcaagggt	atcaccagga	acttctccac	1860
ctgctctccc	agcaccga	cctgccccga	cggccactgc	gatgttgtgg	agaccaaga	1920
catcaacatt	tgcctcagg	actgcctcgg	gggcagcatt	gttgggggac	acgagcctgg	1980
ggagcccccg	gggattaaag	ctggctatgg	cacctgcaac	tgcttccctg	aggaggagaa	2040
gtgcttctgc	gagcccgaa	acatccagga	tcactgtgc	gacgagctgt	gccgcagggt	2100
gatcgcagcc	gctgtcctct	tctccttcat	cgtctcgggt	ctgctgtctg	ccttctgcat	2160
ccactgtac	cacaagtttg	cccacaagcc	acccatctcc	tcagctgaga	tgaccttcgg	2220
gaggccccgc	caggccttcc	cggtcageta	ctcctcttcc	ggtgccccgc	ggccctcgct	2280
ggactccatg	gagaaccagg	tctccgtgga	tgccttcaag	atcctggagg	atccaaagtg	2340
ggaattccct	cggaagaact	tggttcttgg	aaaaactcta	ggagaaggcg	aatttgga	2400
agtggtaag	gcaacggcct	tcacatcgaa	aggcagagca	gggtacacca	cgggtggcgt	2460
gaagatgctg	aaagagaacg	ctccccgag	tgagcttcga	gacctgctgt	cagagtcca	2520
cgtcctgaag	cagggtcaacc	acccacatgt	catcaaatgt	tatggggcct	gcagccagga	2580
tggcccgctc	ctctcctatg	tggagtacgc	caaatacggc	tccctgcggg	gcttctctcg	2640
cagagccgc	aaagtggggc	ctggctacct	gggcagtggg	ggcagccgca	actccagctc	2700
cctggaccac	cggatgagc	gggcctcac	catgggcgac	ctcatctcat	ttgcttggea	2760
gatctcacag	gggatgcagt	atctggccga	gatgaagctc	gttcateggg	acttggcagc	2820
cagaaaacac	ctggtagctg	aggggcggaa	gatgaagatt	tcggatttcg	gcttgtcccg	2880
agatgtttat	gaagaggatt	cctacgtgaa	gaggagccag	ggtcggatc	cagttaaatg	2940
gatggcaatt	gaatcccttt	ttgatcatat	ctacaccacg	caaagtgatg	tatggctttt	3000
tgggtctctg	ctgtgggaga	tcgtgacctc	agggggaaac	ccctatectg	ggattctctc	3060
tgagcggctc	ttcaaccttc	tgaagaccgg	ccaccggatg	gagaggccag	acaactgcag	3120
cagggagatg	taccgcctga	tgttgcaatg	ctggaagcag	gagccggaca	aaaggccggg	3180
gtttgcggac	atcagcaaag	acctggagaa	gatgatgggt	aagaggagag	actacttgga	3240
ccttgccggc	tccactccat	ctgactccct	gatttatgac	gacggcctct	cagaggagga	3300
gacaccgctg	gtggactgta	ataatgcccc	cctccctcga	gccctccctc	ccacatggat	3360
tgaaaacaaa	ctctatggca	tgtcagaccc	gaactggcct	ggagagagtc	ctgtaccact	3420

US 10,450,597 B2

75

76

- continued

```

cacgagagct gatggcacta acactgggtt tccaagatat ccaaatgata gtgtatatgc 3480
taactggatg ctttcacccct cagcggcaaa attaatggac acgtttgata gttaacattt 3540
ctttgtgaaa ggtaaatggac tcacaagggg aagaaacatg ctgagaatgg aaagtctacc 3600
ggccctttct ttgtgaacgt cacattggcc gagecgtgtt cagttcccag gtggcagact 3660
cgttttttgt agtttgtttt aaattccaag gtggttttac ttctgatage cggtgatttt 3720
ccctcctagc agacatgccca caccgggtaa gagctctgag tcttagtggt taagcattcc 3780
tttctcttca gtgccacga gcaccagtg ttggtctgtg tccatcagtg accaccaaca 3840
ttctgtgttc acatgtgtgg gtccaacact tactacctgg tgtatgaaat tggacctgaa 3900
ctgttggtatt tttctagtgt ccgccaaaaa aggcacaaaa atttaaacat gaagcacaca 3960
cacaaaaaag gcagtaggaa aaatgctggc cctgatgacc tgtccttatt cagaatgaga 4020
gactgcgggg ggggcctggg ggtagtgtca atgcccctcc agggctggag ggggaagagg 4080
gccccgagga tgggcctggg ctccagcttc gagatcttga gaatgatttt tttttaatca 4140
tgcaaccttt ccttaggaag acatttggtt ttcacatga ttaagatgat tcttagattt 4200
agcacaaatg agagattcca tgcctctttt actatgtgga tgggtggtatc aggggaagagg 4260
gctcacaaga cacatttgtc ccccgggccc accacatcat cctcacgtgt tcggtactga 4320
gcagccacta cccctgatga gaacagtatg aagaaaaggg gctgttggag tcccagaatt 4380
gctgacagca gaggttttgc tgetgtgaat cccacctgcc accagcctgc agcacacccc 4440
acagccaagt agaggcgaag gcagtggctc atctacctg ttaggagcag gtaggccttg 4500
tactcacttt aatttgaatc ttatcaactt actcataaag ggacaggcta gctagctgtg 4560
ttagaagtag caatgacaaat gaccaaggac tgetacacct ctgattacaa ttctgatgtg 4620
aaaaagatgg tgtttggctc ttatagagcc tgtgtgaaag gccatggat cagetcttcc 4680
tgtgtttgta atttaagtct gctacaagat gtttctgttt cttagattct gacctgact 4740
cataagcttc ttgtcattct tcattgcttg tttgtggtca cagatgcaca acaactctcc 4800
agtcttggtg gggcagcttt tgggaagtct cagcagctct tctggctgtg ttgtcagcac 4860
tgtaacttcg cagaaaaagag tcggattacc aaaacactgc ctgctcttca gaattaaagc 4920
actgatagga cttaaaaatg tctcattcaa atactgtatt ttatatagge atttcacaaa 4980
aacagcaaaa ttgtggcatt ttgtgagccc aaggcttggg tgcgtgtgta atagagcctt 5040
gtggtgtgtg cgcacacacc cagagggaga gtttgaaaaa tgcttattgg acacgtaacc 5100
tggtctaat ttgggtgttt ttccagatcc actgtgataa gttcttttac aaatatctat 5160
agacatggta aacttttgtt ttccagatat gcttaatgat agtcttacta aatgcagaaa 5220
taagaataaa ctttctcaaa ttattaaaaa tgcctacaca gtaagtgtga attgctgcaa 5280
caggtttgtt ctcaggaggg taagaactcc aggtctaaac agctgaccca gtgatgggga 5340
atttatcctt gaccaaatta tccttgacca ataacctaat tgtctattcc tgagttataa 5400
aagtccecat ccttatttagc tctactggaa ttttcataca cgtaaatgca gaagttacta 5460
agtattaagt attactgagt attaagtagt aatctgtcag ttattaaaat ttgtaaaaac 5520
tatttatgaa aggtcattaa accagatcat gttccctttt ttgtaataca ggtgactaag 5580
aaaatcagtt gtgtaataaa aatcatgtat cataaaaaaa aaaaaaaa 5629

```

<210> SEQ ID NO 5

<211> LENGTH: 10

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

US 10,450,597 B2

77

78

-continued

<220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> SEQUENCE: 5

tttttttttt

10

What is claimed herein is:

1. A method of preparing nucleic acids for analysis, the method comprising:

(a) contacting a first nucleic acid template comprising a sequence of a first strand of a double-stranded target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer;

(b) contacting a second nucleic acid template comprising a sequence of a second strand that is complementary to the sequence of the first strand of the double-stranded target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers, wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,

wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and

(c) subjecting the extension product to an amplification reaction comprising successive rounds of polymerase extension of i) a tail primer that comprises a 3' sequence that specifically anneals to the complement of the common sequence and that comprises a 5' tail sequence, and ii) a primer that specifically anneals to the complement of the target-specific hybridization sequence.

2. The method of claim 1, wherein the first strand or the second strand of the double-stranded target nucleic acid is a ribonucleic acid.

3. The method of claim 1, wherein the first strand or the second strand of the double-stranded target nucleic acid is a deoxyribonucleic acid.

4. The method of claim 1, wherein steps (a) and (b) are performed sequentially.

5. The method of claim 1, wherein the first nucleic acid template comprises an extension product resulting from the hybridization and extension of the at least one of the plurality of different primers in step (b).

6. The method of claim 1, wherein the second nucleic acid template comprises an extension product resulting from the hybridization and extension of the target-specific primer in step (a).

7. The method of claim 2, wherein the ribonucleic acid is a messenger RNA encoded from a chromosomal segment that comprises a genetic rearrangement.

8. The method of claim 3, wherein the double-stranded target nucleic acid is a chromosomal segment that comprises a portion of a genetic rearrangement.

9. The method of claim 8, wherein the genetic rearrangement is an inversion, deletion, or translocation.

10. The method of claim 1, further comprising contacting the extension product or amplified extension product with an immobilized oligonucleotide under conditions in which hybridization occurs between the extension product and immobilized oligonucleotide.

11. The method of claim 1, wherein the double-stranded target nucleic acid comprises a target portion having a known sequence and a flanking portion having an unknown sequence.

12. The method of claim 11, wherein different hybridization sequences are complementary to the flanking portion.

13. The method of claim 11, wherein the target-specific hybridization sequence is complementary to the target portion.

14. The method of claim 1, wherein the target-specific primer further comprises, 5' to the target-specific hybridization sequence, at least one of an index sequence, a barcode sequence and an adaptor sequence.

15. The method of claim 1, wherein the common sequence comprises at least one of an index sequence, barcode sequence and an adaptor sequence.

16. The method of claim 14, wherein the adaptor sequence is a cleavable adaptor sequence for immobilizing oligonucleotides in a flow cell.

17. The method of claim 1, wherein the tail primer comprises a nucleic acid sequence identical to the common sequence.

18. The method of claim 1, wherein the 5' tail sequence comprises at least one of an index sequence, a barcode sequence, an adaptor sequence, and a sequencing primer hybridization sequence.

19. The method of claim 1, wherein the primer that specifically anneals to the complement of the target-specific hybridization sequence comprises a sequence identical to the target-specific primer.

20. The method of claim 1, wherein the primer that specifically anneals to the complement of the target-specific hybridization sequence comprises at least one modified internucleoside linkage.

21. The method of claim 1, wherein the target-specific primer comprises at least one modified internucleoside linkage.

22. The method of claim 1 further comprising purifying the extension product prior to (c).

23. The method of claim 1 further comprising purifying the amplified product of (c).

24. The method of claim 1, wherein the plurality of different primers can specifically anneal to sequences comprised by separate genes.

25. The method of claim 1, wherein the plurality of different primers can specifically anneal to different portions

US 10,450,597 B2

79

80

of the first strand or the second strand of the double-stranded target nucleic acid comprised by a single gene.

26. The method of claim 1, wherein the plurality of different primers can specifically anneal to different exons of a gene comprising the first strand or the second strand of the double-stranded target nucleic acid.

27. The method of claim 1, wherein the first strand or the second strand of the double-stranded target nucleic acid is derived from a biological sample.

28. The method of claim 27, wherein the first strand or the second strand of the double-stranded target nucleic acid is a cDNA produced from RNA in the biological sample.

29. The method of claim 1, wherein the double-stranded target nucleic acid is derived from a sample comprising nucleic acids that have been ligated to an oligonucleotide comprising a barcode sequence.

* * * * *

EXHIBIT 2

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

ARCHERDEX, LLC and THE GENERAL)
HOSPITAL CORPORATION d/b/a)
MASSACHUSETTS GENERAL HOSPITAL,)
)

Plaintiffs,)

v.)

C.A. No.

QIAGEN SCIENCES, LLC, QIAGEN) 18-1019-MN
LLC f/k/a QIAGEN, INC., QIAGEN)
BEVERLY, LLC f/k/a QIAGEN)
BEVERLY, INC., QIAGEN)
GAITHERSBURG, INC., QIAGEN)
GMBH, QIAGEN N.V. and JONATHAN)
ARNOLD,)

Defendants.)

REMOTE ORAL AND VIDEOTAPED DEPOSITION OF

DR. MICHAEL METZKER

9th day of February, 2021

VOLUME 1

Pages 1 - 145

Job No. 4454056

1 (Pause.)

2 BY MR. REINES:

3 Q. I'd like to ask you a few questions
4 about the topic of "primers annealing to a
5 sequence."

6 And I know that you've -- you've
7 addressed this topic a few times in this case, but
8 I'd like -- I'd like -- I just want to understand
9 what are the conditions one would have to consider
10 in order to determine whether a primer would anneal
11 to a particular sequence?

12 A. Can you be a little more specific?
13 That's a pretty broad question.

14 Q. Okay. I'll try to be more precise.
15 The -- let's take the '597 patent. The
16 '597 patent sets forth the way to estimate whether
17 a particular primer is going to anneal to a
18 sequence such as length and sequence of the primer,
19 the melting temperature, cell concentration, and
20 things like that.

21 Is that generally familiar to you?

22 A. Yeah. I mean, those are some of the
23 things that you would consider, but you would also
24 have to do actual experimental research to
25 determine whether you have hybridization or not

1 and -- and how good the hybridization is.

2 Q. What are the conditions that you would
3 have to know in order to evaluate whether a
4 particular primer would anneal to a particular
5 sequence?

6 MR. NARAVAGE: Objection, vague.

7 A. Well, I think there are a lot of things
8 that you would have to consider: You can look at
9 the actual sequence, percent GC, the length,
10 complexity of the target.

11 But to really know whether you're
12 getting binding or not, you would design
13 experiments to test the hypothesis whether the
14 binding is correct or incorrect.

15 BY MR. REINES:

16 Q. When you say "binding is correct or
17 incorrect," what do you mean by that?

18 A. Well, whether the sequence as designed
19 would hybridize to the template that it's supposed
20 to bind to. So is it a hundred percent
21 complementary, for instance?

22 Q. In the instance where it's not
23 complementary, would the factors that you expressed
24 regarding GC-content, the length, complexity of the
25 target, the binding, melting temperature and cell

1 concentration, would those all be considerations
2 that could be valued -- evaluated?

3 A. I think they can be evaluated, but I
4 think they're inadequate of addressing the question
5 of whether the binding is correct or not correct.

6 Q. I guess when you say "binding is correct
7 or not correct," I don't know what you mean. Do
8 you mean whether it's successful binding or not?

9 A. No, whether it's binding to its intended
10 target, design of primers with the intent of
11 100 percent complementary binding to its template.

12 Q. Okay. Let -- let's take the case of a
13 primer -- primers can bind to sequences that are
14 not a hundred percent of the mass, correct?

15 A. They can misprime and bind to the
16 wrong -- wrong side of the template.

17 Q. And you consider that to be always
18 unintended to call it "mispriming," that that was
19 something bad?

20 A. Yes. Absolutely. The -- the purpose of
21 designing an oligo is so that it binds to the
22 correct sequence in the template.

23 Q. And whether -- you're using this term
24 "misprime." Whether you can expect mispriming,
25 what are the considerations you would take into

1 account?

2 MR. NARAVAGE: Objection, vague.

3 A. I actually don't understand what you
4 mean by the considerations of mispriming.

5 BY MR. REINES:

6 Q. In the instance of mispriming, there
7 are -- certain primers will misprime to particular
8 sequences, correct, depending on the length, for
9 example?

10 A. Well, it -- the length, GC-content,
11 experimental conditions, annealing temperatures.
12 There's lots of ways to cause mispriming, but
13 that's not what the intent is; it's to actually get
14 correct primer.

15 Q. Have -- have you ever been in a
16 circumstance where -- where you want to evaluate
17 whether a particular experiment is vulnerable to
18 mispriming?

19 A. Sure. I've looked -- I've looked at
20 those kind of effects before.

21 Q. And what do you evaluate to determine if
22 there's vulnerability in a particular experiment to
23 mispriming? What are the considerations that you
24 consider?

25 A. Again, it's primer design, which

1 involves length, GC; you may look at secondary
2 structure within the actual oligonucleotide; you
3 would look at salt conditions; but then you would
4 actually design experiments that actually do real
5 experiments to test whether your considerations are
6 appropriate or insightful.

7 Because, generally, what happens in a
8 test tube is very different than what happens in
9 simulation.

10 So you could design an assay to look at
11 mis -- the -- the percentage of mispriming events
12 based on certain experimental conditions in your --
13 in your experiment.

14 Q. Is mispriming a -- a common issue in
15 your work in -- with targeted assays?

16 A. I -- I think it all depends on the
17 details, but there are clearly -- I
18 have (indiscernible) -- I'm so sorry.

19 There are clearly examples where
20 mispriming occurs. Usually efforts are done to
21 minimize those kind of effects because they are
22 unwanted and undesired.

23 Q. Is mispriming common enough that in
24 evaluating primer extension experiments that you --
25 I mean, do you routinely take into account the

1 potential for mispriming?

2 A. Well, you -- if the assay is designed
3 properly, you can see mispriming occurring, and my
4 approach would be to alter the experimental
5 conditions to minimize that mispriming event,
6 because, again, it's just an undesired, unwanted
7 phenomenon.

8 MR. REINES: I'd like to mark as the
9 Exhibit 2 the '597 patent, please.

10 (Exhibit 2 introduced.)

11 BY MR. REINES:

12 Q. Let me know when you have that, please.

13 A. It just came up and I'm downloading it
14 right now. All right. I have it open.

15 Q. And if you'd turn, please, to
16 paragraph -- actual--"paragraph"--Column 19.

17 A. Okay.

18 Q. Let me know when you have column 19,
19 please.

20 A. Okay. I'm at Column 19.

21 Q. And this is the section of '597 patent
22 that describes what "annealing" is; is that
23 correct?

24 A. I see a sentence around line 51 that
25 says: As used herein, the term "anneal" refers

1 to...

2 Is that what you're referring to?

3 Q. Yes.

4 A. Yes, I see that section.

5 Q. Okay. And then in terms of describing
6 what annealing is in the '597 patent, it refers to:
7 Annealing can involve complementary or
8 substantially complementary nucleic acid strands.

9 Do you see that?

10 A. I do see that.

11 Q. And is that consistent with your
12 experience?

13 A. Yeah. I mean, you -- you can get a
14 primer to anneal to a template sequence that is not
15 100 percent complementary. That's just not -- that
16 doesn't mean that's something that you want the
17 assay -- you don't want the assay to perform in
18 that manner.

19 Q. So in your view, when you're furth --
20 substantially complementary annealing is -- this is
21 what you've called "mispriming" that's unwanted,
22 according to your testimony?

23 MR. NARAVAGE: Objection,
24 mischaracterizes.

25 A. Yeah, I -- I think it depends on the

1 details, but if I -- if I want a primer to bind to
2 a template sequence with a hundred percent
3 complementary and I have a second primer that
4 misprimes, that is an undesired effect.

5 BY MR. REINES:

6 Q. Yeah, that -- that wasn't my question.

7 (Simultaneous conversation.)

8 A. But it -- it --

9 BY MR. REINES:

10 Q. My question is --

11 A. But it's still -- but it's still
12 annealing. There's nothing inconsistent with what
13 the patent is saying and what I am saying.

14 Q. Is the annealing of substantial
15 complementarity always a unwanted event in the
16 nucleic acid sciences?

17 A. I would say for targeted assays, yes.

18 Q. And your explanation for why the '597
19 patent describes its substantial complementarity as
20 part of annealing is just because that's a -- it's
21 kind of a fact of life for people skilled in the
22 art; is that correct?

23 MR. NARAVAGE: Objection,
24 mischaracterizes.

25 A. Well, I think -- I think what's

1 happening here is the applicants are acting as
2 their own lexicographer and they're just trying to
3 provide a definition of what they think "anneal"
4 means.

5 BY MR. REINES:

6 Q. In -- according to teachings of the '597
7 patent, do you believe the patentees are
8 discouraging readers from using primers that only
9 anneal with substantial complementarity as compared
10 to complete complementarity?

11 MR. NARAVAGE: That calls for
12 speculation.

13 A. Could you please repeat the question?

14 BY MR. REINES:

15 Q. Does the '597 patent discourage the use
16 of primers that achieve only substantial
17 complementarity as compared to complete
18 complementarity?

19 A. Can I have a second to look through the
20 patent?

21 Q. Sure. Look at whatever you need to look
22 at to answer throughout the day and tomorrow.

23 A. (Examined exhibit.) Okay.

24 Sorry. If you could just repeat the
25 question, I can now answer.

1 Q. Does the '4 -- '597 patent describe a
2 primer annealing with substantial complementarity
3 as compared to complete complementarity as a
4 negative thing?

5 A. I think so, yes.

6 Q. Now, if you turn to Claim 19.

7 A. (Complied.)

8 Q. If you look down at line 59, do you see
9 where it says: Conditions for annealing between a
10 primer and a nucleic acid tablet may vary based
11 upon length, sequence of the primer?

12 A. I do see that.

13 Q. And do you agree with that?

14 A. Those are some considerations.

15 Q. And conditions for annealing are based
16 on T_m ?

17 A. Correct.

18 Q. And have you calculated T_m in your
19 personal work experience?

20 A. Yes. That's -- that's the melting
21 temperature.

22 Q. Is the melting temperature of a primer,
23 for purposes of evaluating annealing, is that a
24 standard calculation?

25 A. It's -- it's done quite often.

1 Q. And it states, moving over to Column 20,
2 that: The melting temperature can be determined
3 using a number of different algorithms.

4 Do you see those?

5 A. I do see that section.

6 Q. Are you -- are you generally familiar
7 with those algorithms?

8 A. I -- I've used some of these programs
9 myself. They -- they do a number of things
10 including Tm calculations, but the -- that's really
11 kind of the starting point of determining the
12 annealing property of the primer.

13 Q. In terms of the annealing properties of
14 the primer, do you -- have you applied the formula
15 that's set forth in Column 20 at line 15?

16 A. Um, that's a good question. I usually
17 use the software, and most of these software
18 programs use consistent or common formulas for
19 calculating Tm, so possibly.

20 Q. And if you look between line 15 down to
21 line 50, are those descriptions of the
22 considerations to evaluate whether a particular
23 primer will anneal, based on your own expertise?

24 A. Well, I think what this discussion here
25 is about the effect of the primer's Tm and how

1 close below or above the T_m you -- would you expect
2 the primer to bind.

3 The T_m , by definition, is where half of
4 the primer anneals and half of the primer does not
5 anneal, and they are just describing how
6 temperature can influence the amount of primer
7 annealing to its template sequence.

8 Q. And that's consistent with your own
9 understanding, correct?

10 A. That is correct.

11 Q. If you wanted to determine whether a
12 particular primer would anneal to a particular
13 nucleic acid strand at a particular melting
14 temperature, you would use either the software or
15 this formula; is that correct?

16 A. Well, it's a starting point. The actual
17 experimental conditions can influence that T_m .

18 For example, if high salt is used or
19 solvent like DMSO is used or if you use glycerol or
20 a reagent called betaine, many experimental
21 reagents can influence the hybridization properties
22 of the primer.

23 Q. And whether a particular primer will
24 anneal to a particular sequence at particular
25 melting temperatures, the analysis in the '597

1 patent, at Columns 19 and 20, would apply whether
2 the nucleic acid sequence is substantially
3 complementary or completely complementary?

4 A. No, I would not -- I wouldn't
5 characterize it that way. Most of these programs
6 are designed to help select primers that are a
7 hundred percent complementary. They are not
8 designed to design primers that are somewhat
9 complementary. That's not the purpose. The
10 purpose is to design oligonucleotide primers that
11 are complementary to the template sequence of
12 interest.

13 Q. Leaving aside the purposes that you're
14 imputing in the situation, my question to you is:
15 If someone wanted to determine whether a particular
16 primer will anneal to a substantially complementary
17 nucleic acid strand, would the analysis in the '597
18 patent at Claim 20 assist with that process?

19 MR. NARAVAGE: Objection, vague.

20 A. It might assist, but it would be clearly
21 inadequate. It requires more analysis than just
22 those programs listed in the '597 patent.

23 BY MR. REINES:

24 Q. So, you know, clearly, the '597 patent
25 contemplates that annealing involves both

1 complementary and substantially complementary
2 nucleic acid strands, correct?

3 MR. NARAVAGE: Objection,
4 mischaracterizes.

5 A. Well, like I said, I think that portion
6 in Column 19 of the term "anneal," the patentees
7 are just acting as their own lexicographer of how
8 they want to define what the term "anneal" means.

9 BY MR. REINES:

10 Q. Okay. So the -- in the language of the
11 '597 patent, the way the terms have been defined,
12 "annealing" includes both the annealing of a
13 complementary and a substantially complementary
14 nucleic acid; is that correct?

15 MR. NARAVAGE: Objection,
16 mischaracterizes.

17 A. That's a complicated question. Can --
18 I'm sorry. Can you repeat that, please?

19 BY MR. REINES:

20 Q. Sure. And we want the lights on for
21 this one. (Laughter.)

22 A. I -- I know. I'm sorry. I...

23 Q. In terms of the '597 patent, it defines
24 "anneal" to include: A primer that hybridizes to a
25 nucleic acid that's either complementary or

1 substantially complementary, correct? That's
2 line 54.

3 A. That's what they defined "anneal" to be
4 in their patent.

5 Q. Okay. And for purposes of this case,
6 that's how you've understood "anneal," correct?

7 A. For purposes of the '597, I'd have to
8 take that into consideration.

9 Q. Well, let me ask: In applying the
10 concept of annealing in the '597 patent, in this
11 case, you've accepted that it includes both
12 complementary and substantially complementary
13 nucleic acids?

14 A. That's the way they've defined it.

15 Q. Now, in terms of evaluating whether
16 there's successful annealing for either
17 complementary or nucleic acids, you could -- you
18 could apply the analysis in Column 20 of the '597
19 patent, correct?

20 A. Yeah, again, I -- I disagree. I think
21 you can use these software tools as a first go, but
22 it requires much more analysis to determine binding
23 properties of the primer.

24 Q. And what else would you have to do, at
25 a -- at a minimum?

1 A. At a minimum, you would actually do real
2 experimental assays. You could design assays to
3 test the annealing properties of the primer in the
4 conditions that you are interested in.

5 These programs -- these programs --
6 these programs can create misinformation and they
7 can be inaccurate, and there are publications that
8 describe that; and in my own work, I've used these
9 programs to design primers and they don't always do
10 what you expect them to do.

11 Q. What is the kind of misprimer
12 experiences that you have, even if the software
13 tells you that something won't prime --

14 A. I'm sorry.

15 (Simultaneous conversation.)

16 BY MR. REINES:

17 Q. -- nucleic. Do you --

18 A. Something won't prime?

19 Q. Yeah, not a great -- not a great
20 question.

21 In terms of these algorithms, are they
22 both over- and under-inclusive in terms of
23 reporting to the researcher whether a particular
24 primer will anneal to a particular sequence?

25 A. Well, I guess the way I would answer

1 that question--and it's based on my own
2 experience--is I would use the program as a way of
3 designing a primer. I'd test the primer in my
4 assay to determine whether it's doing what I want
5 it to do and the program did an adequate job of
6 selecting a good primer with good annealing
7 properties and things like that.

8 But it's been my experience that these
9 same programs will give you a primer that doesn't
10 work at all, and I have to go back and redesign it
11 and try again.

12 Q. Have you ever used this software to
13 determine whether you can anticipate what you call
14 "mispriming"?

15 A. Sorry. Which software are you referring
16 to?

17 Q. The software in Column 20 of the '597
18 patent.

19 A. Yeah, there are -- there are a couple of
20 primer designed software programs there. I've used
21 Oligo, I've used Primer3, and I believe I've used
22 NetPrimer. I think I've used several of these.
23 They're different programs. And typically, it's --
24 I use these not to assess mispriming. These
25 programs are used for primer design to identify

1 primers that will work in your assay to give you
2 the desired result.

3 For example, if I'm targeting a gene, I
4 want my primers to bind to that gene and simplify
5 that gene so I can characterize it in some way.

6 Q. Now, in instance of a primer that
7 anneals to a substantially complementary nucleic
8 acid strand, will -- will there still be -- can
9 they anneal and there still be -- let me ask it
10 this way:

11 If a primer anneals to a substantially
12 complementary nucleic acid, is it possible that
13 there's no extension possible even if it's
14 annealed?

15 A. There's definitely a possibility. It
16 would not extend in a polymerase reaction, and
17 that's well described in the art, for example, with
18 some allele-specific PCR. That's how that assay
19 works.

20 Q. So what are examples of where you have
21 annealing where -- you know, for example, described
22 in the '597 patent, where you wouldn't have
23 extension?

24 A. And we're talking about polymerase
25 extension.

(Clarification by the Reporter.)

THE WITNESS: Polymerase.

P-O-L-Y-M-E-R-A-S-E.

THE REPORTER: Thank you.

BY MR. REINES:

Q. Correct? That's right?

Do you want me to ask the question again?

A. Well, I was asking for clarification.

Q. Oh. I'm sorry. I must have lost you.

A. When you say "extension," you're talking about polymerase extension. We're talking about --

Q. Yes.

A. -- polymerasing from a primer?

Q. Yes.

And so let me ask you again. What are the circumstances where you have the annealing of the '597 patent but not have an extension?

A. Well, typically, mismatches around the 3' end of the primer doesn't necessarily have to be the last nucleotide on the 3' end, but mismatches around the 3' end of a primer hybridized to a template can result in a nonextension by DNA polymerase.

Q. Does -- as you understand the term

1 "annealing" in the '597 patent and it's Column 19,
2 as you say how they defined it, is it synonymous
3 with hybridization?

4 A. I think, in general, the term
5 "annealing" is synonymous with hybridization. I
6 don't know if they are making it synonymous or not.

7 Q. In terms of how the term "anneal" is
8 used in the '597 patent, does it imply that there's
9 necessarily extension?

10 A. It does not; although, I should actually
11 say -- I should qualify that because I am -- I am
12 looking at Column 19, line 55, and the patent does
13 say: In some embodiments, in the context of an
14 extension reaction, annealing involves the
15 hybridization of a primer to a template such that a
16 primer extension substrate for a template-dependent
17 polymerase enzyme is formed.

18 Q. Okay. So as you understand annealing in
19 the '597 patent, it can be accompanied by extension
20 in some embodiments; is that correct?

21 A. That's what I just read.

22 Q. Now, before, you testified that the
23 melting point of the nucleic acid reflects the
24 point in temperature when half of the primers will
25 hybridize and the other half won't; is that

1 correct?

2 A. Well, per the T_m -- or the melting
3 temperature, but yeah, that's the temperature where
4 half of the primers are bound to the template and
5 half of them are not for that particular sequence,
6 so it's just one sequence that we're talking about.

7 Q. And in terms of what the curve is around
8 that melting point, is that kind of a narrow curve
9 or is it a broad curve in terms of the range of
10 degrees around T_m , which you'll have at least some
11 annealing?

12 A. I haven't looked at those curves in a
13 while. Generally they're fairly sharp, but I think
14 it depends on the experimental conditions.

15 Q. And do you get a broader curve at lower
16 temperatures?

17 A. No. That's not how the -- these are
18 actual -- I mean, there are different ways of
19 calculating T_m . If you're calculating them by
20 these computer programs, they don't look at curves
21 at all. These are all just calculations as we've
22 seen in the -- the patent. Experimentally, you
23 will perform the assay -- the denaturation assay
24 over a range of temperatures, and you'll look where
25 the duplex will melt. In the assay -- and those

1 are the kind of curves that I'm talking about.

2 So they don't melt broadly at low
3 temperature because you started at low temperature
4 and you increase it to high temperature, and that's
5 where you find the inflection point. And that will
6 give you what the real T_m of that hybridization is.

7 Q. What is the -- when you say it's a
8 relatively tight curve, what's the temperature
9 range? I don't know if you can make some
10 generalizations as to what kind of a curve you'd
11 expect around the T_m ?

12 A. Yeah. It's completely dependent on the
13 primer sequence, the primer length, and the
14 experimental conditions. You can -- you can, for
15 the same sequence and same primer and same
16 template, derive different T_m values based on
17 experimental conditions.

18 Q. Why don't we take another break and
19 reconvene at the top of the hour.

20 MR. NARAVAGE: Before we go off the
21 record, can I just take a quick -- it looks like
22 there's another attorney who joins, Keith Toms.
23 And I'd just like to state that for the record that
24 Keith is on, and I'd ask that if any other
25 attendees join, if you could just announce them for

1 the record, please.

2 MR. REINES: Yeah. I mean, I assume
3 they will announce themselves. I'm looking at a
4 different view than you are. I have just the view
5 of the witness.

6 MR. NARAVAGE: Yeah.

7 (Interruption by the reporter.)

8 MR. REINES: I'm hearing that too.
9 Okay. Thank you.

10 VIDEOGRAPHER: 12:49 p.m. We are off
11 the record.

12 (Recess held.)

13 VIDEOGRAPHER: 1:02 p.m. We are on the
14 record.

15 BY MR. REINES:

16 Q. Let me direct you to paragraph 310 of
17 your invalidity report, which has been marked as
18 Exhibit 1.

19 A. Okay.

20 (Interruption by the reporter.)

21 MR. REINES: Nice to hear they work
22 well.

23 A. Okay. I'm at paragraph 310.

24 BY MR. REINES:

25 Q. Do you have it, Dr. Metzker?

1 A. Yes. I have paragraph 310.

2 Q. Paragraph 310. Okay. Before
3 paragraph 310, this is the portion of your report
4 where you're asserting that Fraiture discloses Step
5 1[c]; is that correct?

6 A. That's correct.

7 Q. And part of the requirements of Step
8 1[c] is that each primer of the plurality of
9 different primers involved don't anneal to the same
10 sequence of a double-stranded targeted nucleic acid
11 as any other primer?

12 A. That's what this Step C says.

13 Q. Now, in the plurality of primers in
14 Fraiture, that's a degenerate set of primers that
15 you -- or finding in the APAGene manual?

16 MR. NARAVAGE: Objection, form.

17 A. Yes. Right above Step 1[c], the APAGene
18 Gold manual is listing four different degenerate
19 primers, primer A through primer D.

20 BY MR. REINES:

21 Q. And in terms of whether the different
22 degenerate primers will anneal to the same
23 sequence, show me where in your paragraphs, this
24 310 and 311, for Fraiture that you analyzed that?

25 A. Well, it's more of an explanation of

1 what degenerate primers really are. They're
2 actually just a pool of specific primers, each
3 designed to bind to different sequences. And
4 that's what I say here. I say that their
5 hybridization sequences are different at the 3'
6 end, and they do not anneal to the same sequence of
7 the double-stranded target because they're designed
8 to bind to different sequences. In --

9 Q. Does it say that? Does it say that?

10 A. I'm reading out of paragraph 310, and I
11 read part of the second sentence in paragraph 310.

12 Q. And in terms of -- in terms of whether
13 or not they would anneal as -- you know, for -- due
14 to substantial complementarity, did you perform any
15 experimentation associated with that? Just asking
16 if you performed the experiment.

17 A. I -- I did not perform any experiments.
18 I -- this was -- I'm relying on my extensive
19 experience with PCR, degenerate PCR in the design
20 of primers.

21 Q. Did you perform any analysis of all the
22 different variables that one would have to consider
23 in order to determine whether there would be
24 annealing, such as temperature, experimental
25 conditions, length of the primer?

1 A. Again, I didn't do any experimental data
2 with these degenerate primers.

3 Q. Did you use the software identified in
4 Column 20 of the -- of the patent?

5 A. I -- I didn't. I don't actually think
6 the software in Column 20 really is the appropriate
7 way of using that software.

8 Q. Regardless of whether you think it's
9 appropriate or not, you didn't use any software to
10 determine whether any of the degenerate primers
11 would anneal to the same sequence?

12 A. Again, I didn't do any analysis. This
13 is based on my experience with degenerate PCR, PCR,
14 and primer design.

15 Q. Have you ever used the four primers that
16 are identified in column -- in paragraph 311 of
17 your report?

18 A. I have not personally used these
19 primers.

20 Q. Do you know what the -- the melting
21 point temperature is for any of these four primers?

22 A. I do not.

23 Q. Now, in order to satisfy the Step 1[c]
24 of the '597 patent, did -- this is the combination
25 of Fraiture and the APAGene manual; is that

1 Q. In terms of the APA kit, the Gold kit,
2 do you know what that kit's designed for in terms
3 of what -- what genome or what gene or what
4 sequence?

5 A. My general understanding, it's a Genome
6 Walking Kit that can be used on a number of
7 genomes. At least for the DRT primers, the -- the
8 universal primers, those will be designed based on
9 the fusion that they're looking for.

10 Q. In terms of the DRT primers in the APA
11 kit, are those general purpose for, you know, all
12 sequences and all genomes?

13 A. I don't -- I don't know if it's general
14 purpose or -- or not. I -- I think it can be
15 applied to different genomes, but I don't know how
16 general the purpose is.

17 Q. Do you know if it's specific to any
18 particular genome or sequence?

19 A. That's not my understanding. My
20 understanding is it can be used for a number of
21 genomes.

22 Q. So earlier, you testified that
23 degenerate primer pools would be designed to
24 minimize mispriming. Do you recall that?

25 A. Well, I said -- I think what I said was

1 Did you consider what the actual
2 sequence that's at issue in McCarty to evaluate
3 whether multiple primers would anneal to the same
4 sequence?

5 A. Well, I've reviewed the paper and I've
6 looked at the sequences, and unlike the APAGene,
7 McCarty's degenerate primers are quite simple. I
8 think they have maybe 64 different primer
9 sequences, and, again, I think that's a design to
10 bind to different parts of the genome.

11 Q. So you -- in order to determine whether
12 multiple primers anneal to -- would anneal to the
13 same sequence, do you need to know the sequences in
14 the target nucleic acid?

15 A. I'm sorry. I'm trying to understand
16 your -- can you -- can you ask that for me again,
17 please.

18 Q. In order to determine whether multiple
19 primers would anneal to the same sequence, do you
20 need to understand the sequence of the nucleic acid
21 target?

22 A. Well, I think that's part of -- part of
23 what you need to know. I mean, there are other
24 factors that you need to know as well.

25 Q. And so you considered what the sequence

1 which is --

2 MR. REINES: Let's make the McSwiggen
3 report, please.

4 (Exhibit 5 introduced.)

5 THE WITNESS: I don't have Exhibit 5
6 yet. There it is. Okay. Let me download it.

7 Okay. I have the McSwiggen -- McSwiggen
8 report open.

9 BY MR. REINES:

10 Q. Okay. Yes. If you could look at
11 Exhibit D.

12 THE REPORTER: D as in dog?

13 MR. REINES: D as in dog.

14 THE REPORTER: Thank you.

15 MR. NARAVAGE: I don't think the
16 exhibits are attached to this.

17 MR. REINES: Oh, okay.

18 BY MR. REINES:

19 Q. Well, then -- then let's just look at
20 paragraph 56.

21 A. That's correct. Exhibit D is not
22 attached.

23 56?

24 Q. Yes.

25 A. Okay. I'm there.

1 Q. Do you see where it says in
2 paragraph 56, second to last sentence: On average,
3 there were at least five primer sequences that
4 combined to every MU TIR target sequence with a
5 melting temperature of 60 degrees C or greater?

6 A. I do see that sentence.

7 Q. And did you analyze whether that's
8 correct or not?

9 A. Well, I didn't attempt to repeat this
10 analysis because I -- I think -- this is not --
11 (Simultaneous conversation.)

12 Q. I just asked -- I just asked you whether
13 you did it or not.

14 A. I didn't. I didn't repeat it.

15 Q. And do you have any basis to contest the
16 results that are shown on Table 6?

17 A. Yeah, I do.

18 Q. And what's your contention regarding
19 that?

20 A. Well, I think I address the McSwiggen
21 report in my reply report, but sure, we've done
22 this analysis on computational studies, and this is
23 not the right way to do it.

24 The analysis of TMs without any
25 experimental data is -- is very inadequate and can

1 lead to false information.

2 So I don't -- I don't agree that this
3 actually represents mispriming as the McSwiggen
4 report says.

5 Q. Did you attempt to perform experiments
6 that would show whether the computational analysis
7 presented by McSwiggen would be accurate or not?

8 A. I did not do any computational analysis.

9 Q. My question wasn't whether you did any
10 computational analysis. My question was whether
11 you did any experimental analysis that might
12 disprove the computational analysis of McSwiggen
13 that you're criticizing?

14 A. I'm sorry. I misheard your question.

15 No, I did not do any experimental
16 analysis.

17 Q. When a primer differs from a sequence by
18 one base, what's the likelihood of hybridization or
19 annealing?

20 A. I think it is completely dependent on
21 the experimental conditions.

22 Q. Do you have an opinion as to whether
23 the -- let me say this: In terms of McCarty, do
24 you -- do you believe the Mu-seq is not a
25 target-specific hybridization sequence?

EXHIBIT 3

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE**

ARCHERDX, INC. and THE GENERAL
HOSPITAL CORPORATION d/b/a
MASSACHUSETTS GENERAL
HOSPITAL,

Plaintiffs,

v.

QIAGEN SCIENCES, LLC, QIAGEN LLC
f/k/a QIAGEN, INC., QIAGEN BEVERLY,
INC., QIAGEN GAITHERSBURG, INC.,
QIAGEN GMBH, QIAGEN N.V., and
JONATHAN ARNOLD,

Defendants.

Civil Action No. 18-cv-01019-MN-CJB

**HIGHLY CONFIDENTIAL –
ATTORNEYS’ EYES ONLY**

**EXPERT REPORT OF DR. MICHAEL METZKER REGARDING
THE INVALIDITY OF U.S. PATENT NOS. 10,017,810 AND 10,450,597**

TABLE OF CONTENTS

I.	QUALIFICATIONS/COMPENSATION/MATERIALS REVIEWED	1
A.	Education/Training/Expertise	1
B.	Compensation	2
C.	Materials Reviewed	2
II.	TECHNICAL BACKGROUND.....	3
A.	DNA Basics	3
B.	PCR.....	6
C.	DNA sequencing technologies.....	7
D.	Target Enrichment	9
III.	APPLICABLE STANDARDS	13
A.	The Pertinent Art and One of Ordinary Skill in the Art	14
B.	Anticipation Standards.....	15
C.	Obviousness Standards	15
D.	Written Description and Indefiniteness Standards.....	17
E.	The Meaning of Claim Terms and Scope of the Asserted Claims to One of Ordinary Skill in the Art	18
IV.	THE PATENTS IN SUIT AND ASSERTED CLAIMS	20
A.	Patent Priority of the Asserted Claims of the Patents-in-Suit.....	21
V.	U.S. PATENT NO. 10,017,810 (Claims 16-20).....	21
A.	The Asserted Claims Are Invalid for Lack of Written Description and Indefiniteness....	21
B.	The Asserted Claims Are Invalid Because They Are Anticipated or Obvious.....	28
1.	Amsterdam Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	39
2.	Siebert Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art.....	60
3.	Markham Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art.....	76
4.	Chenchik Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art.....	121
VI.	U.S. PATENT NO. 10,450,597 (Claims 1, 3-5, 10, 11, 13-27, and 29).....	155
A.	The Asserted Claims Are Invalid for Lack of Written Description and Indefiniteness..	157
B.	The Asserted Claims Are Invalid Because They Are Anticipated or Obvious.....	159
1.	Fraiture Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art.....	159

2. McCarty Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art 194

3. Morley Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art 225

4. Martin-Harris Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art 260

5. The APAGene Manual Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art 293

6. The '810 and '828 Patents Anticipate the Asserted Claims or Render Them Obvious in Combination with Other Prior Art 328

VII. TRADE SECRET 348

VI. U.S. PATENT NO. 10,450,597 (Claims 1, 3-5, 10, 11, 13-27, and 29)

293. The claimed method of the '597 Patent, like that of the '810 Patent, claims the well-known prior art technique of chromosome walking or genome walking, adapted for next generation sequencing, which I described above.²⁴⁹ The claimed method of the '597 Patent involves using one primer (a “target-specific” primer) that binds to a double-stranded target nucleic acid to carry out a single primer extension (“SPE”) reaction. The target-specific primer contains a “target-specific hybridization sequence” that anneals to the target. A second round of SPE is performed using a plurality of different primers that share a common sequence 5' to different hybridization sequences. Through this process, an “extension product” is generated that is then amplified using PCR. These PCR reactions are performed using a “tail” primer that includes a 3' sequence that anneals to the complement of the common sequence on the plurality of primers, and a 5' tail sequence, and then in the opposite direction a primer that anneals to the complement of the target-specific hybridization sequence.

294. This method, like the method of the '810 Patent, is intended to allow “walking” along a genome from a sequence that is known to a sequence that is unknown. The '810 Patent Asserted Claims require a universal adaptor that is ligated to all the fragments at the “unknown” end of the target, to permit capture of those unknown sequences and therefore “walking” from the known target nucleotide sequence to the unknown target nucleotide sequence. The '597 Patent Asserted Claims carry out the genome walking method with the claimed “plurality” of primers that

²⁴⁹ See, e.g., Ex. PP, QIA-00689179-00689181, Martin-Harris *et al.*, *Gene walking using sequential hybrid primer polymerase chain reaction*, ANALYTICAL CHEMISTRY 399:308-310 (2010) (“Martin-Harris”); Ex. QQ, QIA-00689151-00689159, Fraiture *et al.*, *An innovative and integrated approach based on DNA walking to identify unauthorized GMOs*, FOOD CHEMISTRY 147:60-69 (2014) (“Fraiture”); QIA-00689264-00689363, Ex. RR, Morley, *Method of DNA amplification*, UNITED STATES PATENT APPLICATION NO. US 2010/0248310 (2010) (“Morley”).

would be understood by one of ordinary skill in the art at the relevant time to refer to “degenerate primers” or “degenerate random primers.”

295. “Degenerate primers” include sequences that are not fixed, but are generated randomly, so that a pool of degenerate primers will include primers with a variety of sequences that they can bind to sequences throughout the genome. The idea is that, instead of a universal adaptor ligated to the unknown end of the target sequence, there will be a pool of primers, one of which can bind to the unknown region of the target. These primers were commonly used for sequencing unknown regions flanking a known target sequence, such as in gene walking.²⁵⁰ These degenerate primers were generally paired with target-specific primers.²⁵¹

296. Otherwise, if the plurality of primers were not degenerate or random, but instead had a specific sequence, the claimed method of the ’597 Patent would be essentially the same as standard PCR, as described about in the PCR section. My opinion in this regard is supported by the testimony of Drs. Iafrate and Le, both of whom testified that they had not invented standard PCR, and also testified that their method was distinguishable from standard PCR in which the two opposing primers anneal to known sequence.²⁵²

297. As I discussed above, NGS platforms require the addition of platform-specific sequences to the nucleic acid sample that is going to be sequenced on the platform. Also as I discussed above, one of ordinary skill in the art was aware that there were a variety of methods available to attach platform-specific sequences to a nucleic acid sample. One approach, well

²⁵⁰ See, e.g., SS, QIA-00689182-00689192, McCarty *et al.*, *Mu-seq: Sequence-based mapping and identification of transposon induced mutations*, PLOS ONE 8:e77172 pp. 1-11 (2013) (“McCarty”); see also Morley; Ex. TT, QIA-00689102-00689115 Bio S&T Inc., *APAgene™ GOLD-RT genome walking kits* (2010) (“APAgene Manual”); Fraiture.

²⁵¹ See, e.g., *id.*

²⁵² See Ex. E (Iafrate Dep. Tr.) at 69:20-71:7, 72:25-73:6; Ex. F (Le Dep. Tr.) at 108:20-109:9.

known in the art, was to attach these platform-specific sequences by carrying out PCR reactions with primers that included the platform-specific sequences.²⁵³

Like the '810 Patent, the '597 Patent claims certain of these well-known methods as limitations in the dependent claims. These methods include the addition of Illumina adaptor sequences or index sequences to the target nucleic acid sequences.²⁵⁴

298. As I discussed above, multiple prior research groups had already adapted the well-known chromosome walking or genome walking methods to NGS platforms including the Illumina NGS platform and the 454 NGS platform, using adaptors, PCR primers, and combinations of adaptors and PCR primers.

A. The Asserted Claims Are Invalid for Lack of Written Description and Indefiniteness

299. Under Plaintiffs' overbroad reading of the '597 Patent Asserted Claims, as set out in their infringement contentions, the claims are invalid under 35 U.S.C. § 112. Under this overbroad reading, the '597 Patent Asserted Claims lack written description in several respects.

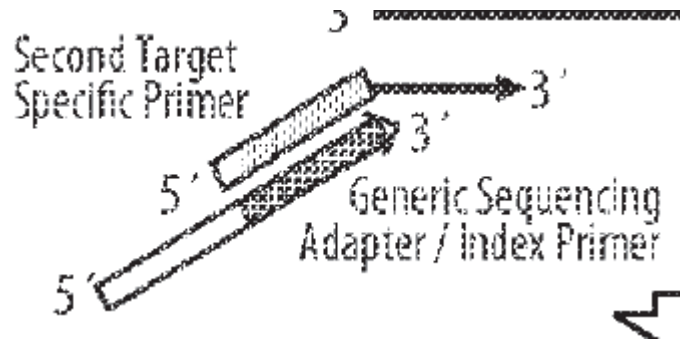
300. Plaintiffs' infringement contentions depend on reading a universal adaptor primer in the QIAGEN Accused Products as a "target-specific primer." According to this reading, the "target" is the universal adaptor. I disagree.

301. First, one of ordinary skill in the art would not have understood from the '597 Patent that the nucleic acid template, or the target specific section of the nucleic acid template, is a universal adaptor sequence ligated to the nucleic acid fragments from the sample. Similarly, the skilled artisan would not have understood from the '597 Patent that the target specific primer is a universal primer that anneals to all universal adaptors that have been ligated to the nucleic acid

²⁵³ See, e.g., McCarty.

²⁵⁴ See Ex. UU, QIA-00689087-00689-092, Illumina, *Genomic sequencing – Data sheet: Sequencing* (2010) ("Illumina Genomic Sequencing Data Sheet"); see also McCarty.

fragments from the sample. In fact, the '597 Patent specification contradicts this interpretation, as shown in the following figure, which expressly identifies a primer binding to a universal adaptor sequence as a “Generic Sequencing Adapter / Index Primer,” in contrast to a “Second Target Specific Primer,” shown binding to the target.²⁵⁵



Plaintiffs’ application of the '597 Patent Asserted Claims in its infringement contentions is also inconsistent with the Court’s Claim Construction for the '597 Patent.²⁵⁶

302. Under Plaintiffs’ reading, the '597 Patent Asserted Claims are also indefinite under 35 U.S. C. § 112. The claims are indefinite if step (b) of claim 1 of the '597 Patent is to be performed before step (a), as Plaintiffs appear to be asserting.²⁵⁷ The claim does not provide one of ordinary skill in the art with reasonable certainty about the scope of the claim. The skilled artisan would have understood that under Plaintiffs’ reading, there is no antecedent basis for the phrase “the first strand of the double-stranded target nucleic acid” in step (b). Similarly, one of ordinary skill in the art would have understood that under Plaintiffs’ reading, there is no clear antecedent basis in step (b) for “the same sequence of the double stranded target nucleic acid.” A first nucleic acid sequence is identified in step (a), and a second nucleic acid sequence is identified in step (b), but “the same sequence” does not identify which of those sequences it is referring to.

²⁵⁵ See '597 Patent at Figure 1C; see also Ex. E (Iafrate Dep. Tr.) at 99:16-102:4.

²⁵⁶ See D.I. 254, *Memorandum Order* at pp. 1-2.

²⁵⁷ See Ex. I (July 16, 2020 Infringement Contentions) at Ex. B p. 31-34.

B. The Asserted Claims Are Invalid Because They Are Anticipated or Obvious

303. As I set out in detail below, the method of the 597 Patent Asserted Claims for preparing nucleic acids for analysis were invented by multiple independent research groups, beginning years before the earliest priority date of the '597 Patent, and were also previously published by inventors of the '597 Patent. These prior art references include Martin-Harris (2010), Morley (2010), Fraiture (2013), McCarty (2013), and the '461 Patent Pub. (2013).²⁵⁸

1. Fraiture Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art

(a) 1[pre]. A method of preparing nucleic acids for analysis, the method comprising:

304. Although I do not believe that the preamble is a limitation, to the extent that the preamble is a limitation, Fraiture discloses the preamble. For example, Fraiture states that the goal of the method it sets out is to prepare nucleic acids for analysis to identify genetic markers to determine whether an organism has been genetically modified. See Fraiture at 63 (“

The aim of the present study is to supply an integrated approach to identify unauthorised GMOs: A first real-time PCR screening allows the detection of the terminator 35S (t35S) of the pCAMBIA family vectors to indicate the potential presence of unauthorised GMOs in food matrices (Fig. 3). ***Then, an appropriate DNA walking method, anchored on the sequence used for the screening followed by two semi-nested PCRs to identify the junction, confirms the GMO presence.***²⁵⁹

Fraiture also describes their method of preparing nucleic acids for analysis having substantial advantages to current methods.

With regard to the previous articles describing methods characterising the junction sequences of GMOs, ***the present DNA walking approach possesses the substantial advantage to be easily implementable in semi-routine use thanks to the simplicity of a method exclusively based on PCR.***

²⁵⁸ See QIA-00689364-00689415, Iafrate *et al.*, *Methods for determining a nucleotide sequence*, UNITED STATES PATENT APPLICATION PUBLICATION NO. US 2013/0303461 (2013) (“’461 Patent Pub.”).

²⁵⁹ See Fraiture at 63 (emphasis added).

*Kit (BIO S&T, Montréal, Canada).*²⁷¹

The sequences of the DRT primers, as listed in the APAGene™ GOLD manual, demonstrate that the DRT primers share a common sequence that is 5' to different hybridization sequences (see common sequence highlighted in yellow below).²⁷²

DRT Primer	Sequences
DRT Primer A:	5'-GAACACGCGTCGTTTACCTCCXXXXGXXXXTAGT-3'
DRT Primer B:	5'-GAACACGCGTCGTTTACCTCCXXXXXTAGT-3'
DRT Primer C:	5'-GAACACGCGTCGTTTACCTCCXXXXXTCAT-3'
DRT Primer D:	5'-GAACACGCGTCGTTTACCTCCXXXXGXXXXCCTG-3'

- (d) **1[c]. wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,**

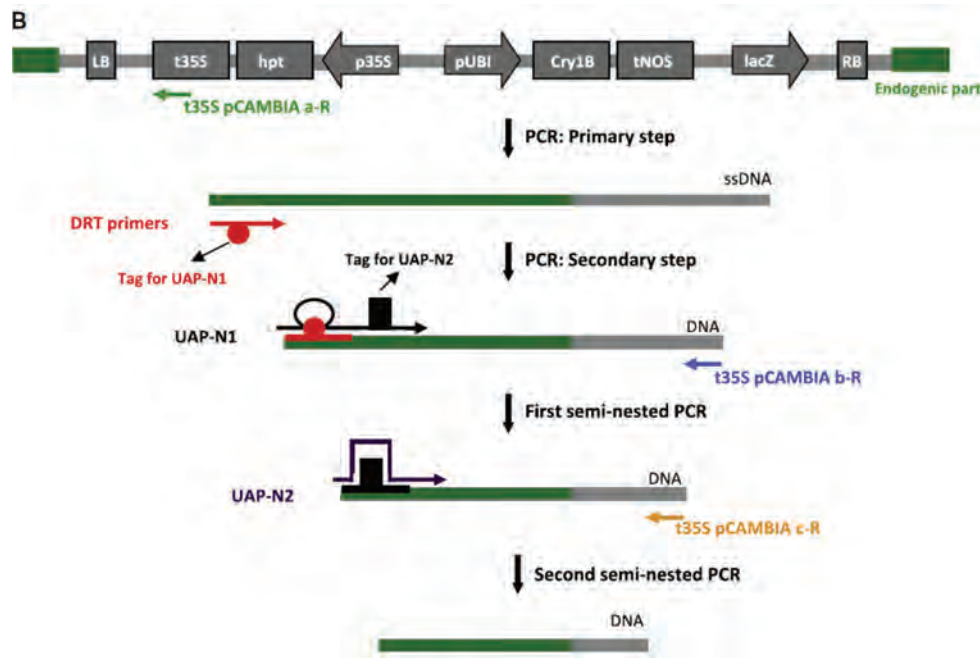
310. Fraiture discloses this limitation, taking into account the parties' agreement that the phrase "the same sequence" is to be interpreted as "the identical sequence."²⁷³ One of ordinary skill in the art at the relevant time would have understood that because the DRT primers are degenerate at their 3' ends, their hybridization sequences are different at the 3' ends, and do not anneal to the same sequence of the double-stranded target as any other degenerate DRT primer. Fraiture discloses the use of different, degenerate hybridization sequences (contained within the DRT primers) in Figure 1B (reproduced below) through the use of the plural "DRT primers."²⁷⁴

²⁷¹ See *id.* at 65 (emphases added).

²⁷² See APAGene Manual at 3, stating that "[t]he sequences of the DRT primers are as follows. Please note that the letter X symbolizes degenerate and/or cleavable nucleotides."

²⁷³ See D.I. 254, *Memorandum Order* at p. 1.

²⁷⁴ See Fraiture at Figure 1.



311. As I discuss above, Fraiture discloses that the DRT primers were provided as part of the APAGene Gold Genome Walking Kit.

2.6.1 DNA walking and double semi-nested PCR

. . . ***Degenerated random tagging (DRT)*** and Universal tagging primers (UAP-N1 and N2) ***were provided by APAGene™ GOLD Genome Walking Kit (BIO S&T, Montréal, Canada).***²⁷⁵

The APAGene manual for the DRT primers lists the sequences of the DRT primers, demonstrating that the DRT primers have different 3' ends and that each DRT primer does not anneal to the same sequence of the double-stranded target nucleic acid as any other DRT primer (the degenerate 3' ends are highlighted in magenta below).²⁷⁶

²⁷⁵ See *id.* at 65 (emphases added).

²⁷⁶ See APAGene Manual at 3, stating that “[t]he sequences of the DRT primers are as follows. Please note that the letter X symbolizes degenerate and/or cleavable nucleotides.”

DRT Primer	Sequences
DRT Primer A:	5'-GAACACGCGTCGTTTACCTCCXXXXGXXXXTAGT-3'
DRT Primer B:	5'-GAACACGCGTCGTTTACCTCCXXXXXXTAGT-3'
DRT Primer C:	5'-GAACACGCGTCGTTTACCTCCXXXXXXTCAT-3'
DRT Primer D:	5'-GAACACGCGTCGTTTACCTCCXXXXXGXXXXCCTG-3'

- (e) 1[d]. wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and,

312. Taking into account the Court’s construction of the term “a sequence characteristic of the target-specific primer” to mean “a sequence from one of the strands of the single target nucleic acid recited in step (a),”²⁷⁷ Fraiture discloses this limitation. Fraiture’s method involves a what Fraiture describes as a “two step PCR reaction” that takes place in one tube, in which the target-specific primer (t35S pCAMBIA a-R primer, shown in green) is extended in an SPE reaction, followed by a second extension reaction using one of the DRT primers.

First, a two-step PCR taking place in one tube is carried out. In the first step, single strand DNA (ssDNA) fragments are produced by a single primer extension reaction using t35S pCAMBIA a-R primer (green). This reaction is repeated four times in four individual tubes. ***In the second step, four different DRT primers (A–D) (red) are immediately added individually to the four reaction tubes.***²⁷⁸

313. The product of these two reactions contains a sequence characteristic of the target-specific primer, in that it contains a “green” sequence characteristic of the “green” primer t35S pCAMBIA a-R.²⁷⁹ The product of these two reactions also contains a sequence characteristic of at least one of the plurality of different primers, in that it contains a “red” sequence characteristic

²⁷⁷ See D.I. 254, *Memorandum Order* at p. 2.

²⁷⁸ See Fraiture at Figure 1 legend (emphases added).

²⁷⁹ See *id.* at Figure 1A and 1B.

active group of Mu transposons. In addition to conferring specificity, the nested Mu-seq primer incorporates a 4-base barcode and a segment of the Illumina sequencing adaptor.”); *id.* at 9 (“To incorporate custom, Mu-specific, multiplexed Illumina sequencing adaptors, 100 ng of PCR I amplicons were subjected to a second round of PCR using a nested Mu-seq primer that includes the 3’-end of the Illumina sequencing adaptor and an Illumina A2-tiB adaptor primer. A series of 64, Mu-seq adaptor primers were synthesized that each incorporate a 4-base barcode between a 20-base segment of the Illumina sequencing adaptor and a nested Mu-TIR primer sequence as follows:

MuSeq-NNNN: 5’-
ACACGACGCTCTTCCGATCTNNNNCBCTCTTCKTCYATAATGGCAAT-3’, where the
barcode is indicated by N’s and the Mu-TIR specific sequence is underlined”).

- (d) **1[c]. wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,**

388. McCarty discloses this limitation, taking into account the parties' agreement that the phrase “the same sequence” is to be interpreted as “the identical sequence.” A person of ordinary skill in the art at the relevant time would have understood that because the Mu-seq primers are degenerate at their 3' ends, their hybridization sequences are different at the 3' ends, and do not anneal to the same sequence of the double stranded target as any other degenerate Mu-seq primer.

389. McCarty describes that each of the nested Mu-seq primers comprise different hybridization sequences with different 3' ends. McCarty describes that the nested Mu-seq primer has the following hybridization sequence (Mu-TIR primer sequence): CBCTCTTCKTCYATAATGGCAAT. McCarty at 9 (“A series of 64, Mu-seq adaptor primers were synthesized that each incorporate a 4-base barcode between a 20-base segment of the Illumina sequencing adaptor and a nested Mu-TIR primer sequence as follows: MuSeq-NNNN: 5’-

ACACGACGCTCTTCCGATCTNNNNCBCTCTTCKTCYATAATGGCAAT-3', where the barcode is indicated by N's and the Mu-TIR specific sequence is underlined.'). The disclosed sequences includes the nomenclature B, K, and Y for nucleotide bases, each of these labels referring to one of at least two different types of bases. A person of ordinary skill in the art would understand that this nomenclature for incompletely specified bases in nucleic acid sequences was classified by the Nomenclature Committee of the International Union of Biochemistry in 1985. See Cornish-Bowden²⁹⁸ at 3021. The nomenclature "B" means that the nucleotide in that position may be any of the three bases guanine or cytosine or thymine. See Cornish-Bowden at 3025. The "K" nomenclature refers to either of the bases guanine or thymine. *Id.* The "Y" nomenclature refers to either thymine or cytosine. *Id.* at 3024. Thus, each of the Mu-seq adaptor primers comprise different hybridization sequences as each primer has different nucleotide bases at any of the positions signified by B, K, or Y.

390. As each of the primers have different nucleotide bases at each of these positions, the hybridization sequences for these primers have different 3' ends. Given that each primer has a different 3' end, each primer does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers.

(e) **1[d]. wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and,**

391. Taking into account the Court's construction of the term "a sequence characteristic of the target-specific primer" to mean "a sequence from one of the strands of the single target nucleic acid recited in step (a)," McCarty discloses this limitation.

²⁹⁸ See Ex. ZZ, QIA-00689116, Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic acids research*, 13(9), 3021 ("Cornish-Bowden").

461. In Example 3, Morley teaches that a PCR reaction was performed using a degenerate primer with a tag sequence of 18 bases (common sequence), 5' to 5 degenerate bases (different hybridization sequences), followed by 5 fixed bases at the 3' end. Morley at ¶ [0250] ("Gene walking along three genes, APC, BRCA1 and myocillin, was performed using 50 ng of a gene-specific forward primer, 50ng of one of a variety of degenerate reverse primers, and 50 ng of a reverse tag primer. The degenerate reverse primers had 4-6 random normal residues at the 3' end, followed by 3-6 degenerate residues, followed by a random tag sequence of 12-18, usually 18, normal residues. The most commonly used degenerate primer had 5 fixed bases at the 3' end followed by 5 degenerate bases, followed by a tag sequence of 18 bases (5'TGCTAGGATCCAAG GNNNNNATTCG3' (SEQID NO: 1)). The reverse tag primer had the same sequence as the tag on the degenerate reverse primer").

(d) **1[c]. wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,**

462. Morley discloses this limitation, taking into account the parties' agreement that the phrase "the same sequence" is to be interpreted as "the identical sequence. For example, Morley discloses degenerate primers wherein the different hybridization sequences have different 3' ends and wherein the different primers anneal to different sequences of the target nucleic acid.

463. In Example 3, Morley discloses the use of degenerate primers each having 3-6 degenerate residues and most commonly 5 degenerate residues, which represents a plurality of 1,024 different primer sequences. See Morley at ¶ [0250]. Because each degenerate primer has degenerate residues 3' of their common sequence, the degenerate primers therefore have different 3' ends, and each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid.

464. Morley also shows an exemplary degenerate primer in Figure 13, the portion highlighted in magenta (NNNNN), which is 3' of the common sequence highlighted in yellow, being different for each primer of the plurality of different primers.

Gene walking with degenerate primer

Degenerate Primer

TGCTAGGATCCAAGGNNNNNATTCG

- (e) **1[d]. wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and,**

465. Morley discloses this limitation. Taking into account the Court's construction of the term "a sequence characteristic of the target-specific primer" to mean "a sequence from one of the strands of the single target nucleic acid recited in step (a)," Morley discloses this limitation.

466. For example, Morley discloses that an extension product is generated characteristic of the gene specific primer and at least one of the degenerate primers. *See, e.g.*, Morley at Figure 4; ¶ [0250] (Example 3) ("Gene walking along three genes, APC, BRCA1 and myocillin, was performed using 50 ng of a gene-specific forward primer, 50ng of one of a variety of degenerate reverse primers, and 50 ng of a reverse tag primer. The degenerate reverse primers had 4-6 random normal residues at the 3' end, followed by 3-6 degenerate residues, followed by a random tag sequence of 12-18, usually 18, normal residues. The most commonly used degenerate primer had 5 fixed bases at the 3' end followed by 5 degenerate bases, followed by a tag sequence of 18 bases (5'-TGCTAGGATCCAAG GNNNNNATTCG-3' (SEQID NO: 1). The reverse tag primer had the same sequence as the tag on the degenerate reverse primer. Five PCR cycles, with annealing at 35° C. for five minutes, were followed by 15 cycles with annealing at 55° C. for 3 minutes."). This disclosed primary PCR results in the generation of an extension product that contains both a

sequence that is characteristic of the target-specific primer (“gene-specific forward primer”) and a sequence that is characteristic of the at least one of the plurality of different primers (“degenerate reverse primers”).

- (f) **1[e]. (c) subjecting the extension product to an amplification reaction comprising successive rounds of polymerase extension of i) a tail primer that comprises a 3’ sequence that specifically anneals to the complement of the common sequence and that comprises a 5’ tail sequence, and ii) a primer that specifically anneals to the complement of the target-specific hybridization sequence.**

467. Morley discloses this limitation.

468. As shown in Fig. 4 of Morley, below, the disclosed method involves subjecting the extension product created using the forward gene-specific primer and the reverse degenerate primers to a PCR amplification reaction comprising successive rounds of extension using a tail primer (“tag1”) that specifically anneals to the complement of the common sequence of the degenerate reverse primers and that also comprises a 5’ tail sequence (“tag2”), and also using another gene-specific primer that specifically anneals to the complement of the target-specific hybridization sequence.

that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers. The “degenerate primer-tag1” on the right of this figure (highlighted in yellow) indicates a plurality of different primers that share a common sequence, the “tag1” sequence, that is 5' to different hybridization sequences under hybridization conditions. See Martin-Harris at 309 (“**A primary PCR**, using a gene-specific primer **and a degenerate primer**, is followed by two or three sequential rounds of PCR, each of which uses a hybrid primer. The 3' end of the hybrid primer binds to a target sequence generated in the preceding round of PCR, and the 5' end of the hybrid primer is a tag sequence that generates a target sequence for the next round of PCR,” “[t]he various degenerate reverse primers had 4-6 random normal residues at the 3' end, followed by 3-6 degenerate residues, followed by a random tag sequence of 12-18 (usually 15) normal residues”) (emphases added), and Table 1.

- (d) **1[c]. wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,**

554. Martin-Harris teaches this limitation, taking into account the parties' agreement that the phrase “the same sequence” is to be interpreted as “the identical sequence.”³⁰⁷

555. For example, Martin-Harris discloses primers “wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers”

³⁰⁷ See D.I. 254, *Memorandum Order* at p. 1.

556. Specifically, Martin-Harris discloses four degenerate primer sequences in Table 1, each having an “NNNNN” sequence (highlighted in yellow), which represents a plurality of 1,024 different primer sequences. *See* Martin-Harris at 309 (Table 1):

Gene walking A	TGCTAGGATCCAAGGNNNNNATTTCG
Gene walking T	TGCTAGGATCCAAGGNNNNNTTTTCG
Gene walking C	TGCTAGGATCCAAGGNNNNNCTTTCG
Gene walking G	TGCTAGGATCCAAGGNNNNNGTTTCG

557. As shown by the table, each of the plurality of degenerate primers have different 3' ends, and each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as other primers of the plurality of different primers.

558. This example also shows that the four degenerate primer sequences have the same common sequence (called “tag1” and highlighted in yellow) that is 5' to the different hybridization sequences. *See* Martin-Harris at 309 (Table 1).

Gene walking A	TGCTAGGATCCAAGGNNNNNATTTCG
Gene walking T	TGCTAGGATCCAAGGNNNNNTTTTCG
Gene walking C	TGCTAGGATCCAAGGNNNNNCTTTCG
Gene walking G	TGCTAGGATCCAAGGNNNNNGTTTCG

- (e) **1[d]. wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and,**

559. Taking into account the Court’s construction of the term “a sequence characteristic of the target-specific primer” to mean “a sequence from one of the strands of the single target nucleic acid recited in step (a),”³⁰⁸ Martin-Harris discloses this limitation.

560. For example, Martin-Harris teaches that “wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific

³⁰⁸ *See* D.I. 254, *Memorandum Order* at p. 2.

tubes. Step 2 immediately follows in which four different DRT primers, along with a 5X DRT binding mixture, are added individually to the four reaction tubes”).

642. Further, the APAgene Manual discloses and the System practices that each of the DRT primers (which constitute a plurality of different primers) comprise a common sequence (GAACACGCGTCGTTTACCT) that is 5’ to different hybridization sequences. See APAgene Manual at 3 (stating that “The sequences of the DRT primers are as follows. Please note that the letter X symbolizes degenerate and/or cleavable nucleotides”) (Table of DRT Primer Sequences reproduced below with common sequence highlighted in yellow and hybridization sequences highlighted in magenta).

DRT Primer	Sequences
DRT Primer A:	5’-GAACACGCGTCGTTTACCTCCXXXXGXXXXTAGT-3’
DRT Primer B:	5’-GAACACGCGTCGTTTACCTCCXXXXXXTAGT-3’
DRT Primer C:	5’-GAACACGCGTCGTTTACCTCCXXXXXXTCAT-3’
DRT Primer D:	5’-GAACACGCGTCGTTTACCTCCXXXXGXXXXCCTG-3’

- (d) **1[c]. wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,**

643. The APAgene Manual teaches and the System practices this limitation, taking into account the parties’ agreement that the phrase “the same sequence” is to be interpreted as “the identical sequence.”³¹⁶ One of ordinary skill in the art at the relevant time would have understood that because the DRT primers are degenerate at their 3’ ends, their hybridization sequences are different at the 3’ ends, and do not anneal to the same sequence of the double-stranded target as any other degenerate DRT primer.

³¹⁶ See D.I. 254, Memorandum Order at p. 1.

644. For example, the APAGene manual for the DRT primers lists the sequences of the DRT primers, which demonstrate that the DRT primers have different hybridization sequences with different 3' ends and that each DRT primer does not anneal to the same sequence of the double-stranded target nucleic acid as any other DRT primer (see different hybridization sequences with different 3' ends for annealing to different sequences highlighted below). See APAGene Manual at 3:

DRT Primer	Sequences
DRT Primer A:	5'-GAACACGCGTCGTTTACCTCCXXXXGXXXXTAGT-3'
DRT Primer B:	5'-GAACACGCGTCGTTTACCTCCXXXXXXTAGT-3'
DRT Primer C:	5'-GAACACGCGTCGTTTACCTCCXXXXXXTCAT-3'
DRT Primer D:	5'-GAACACGCGTCGTTTACCTCCXXXXXGXXXXCCTG-3'

645. Further, the APAGene Manual discloses that the letter X symbolizes degenerate and/or cleavable nucleotides. See APAGene Manual at 3 (“The sequences of the DRT primers are as follows. Please note that the letter X symbolizes degenerate and/or cleavable nucleotides”).

(e) **1[d]. wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and,**

646. Taking into account the Court’s construction of the term “a sequence characteristic of the target-specific primer” to mean “a sequence from one of the strands of the single target nucleic acid recited in step (a),”³¹⁷ the APAGene Manual and the System practices this limitation.

647. The APAGene Manual teaches, and the System practices, that Step 1 of Primary PCR (limitation (a) of Claim 1) results in amplification of single-stranded target DNA fragments by linear amplification using GSPa. See APAGene Manual at 7 (“2.2.2 Step 1 of Primary PCR: This step will amplify single-stranded targeted DNA fragments by linear amplification using

³¹⁷ See D.I. 254, Memorandum Order at p. 2.

multiple samples can be of different origins, *e.g.* from different tissues and/or different subjects. In embodiments of the methods described herein, multiplex applications can include determining the nucleotide sequence contiguous to one or more known target nucleotide sequences in multiple samples in one sequencing reaction or sequencing run. The multiple samples can be of different origins, *e.g.* from different tissues and/or different subjects. In such embodiments, the universal oligonucleotide tail-adaptor can further comprise a barcode portion. In some embodiments, a universal oligonucleotide tail-adaptor with a unique barcode portion can be added to each sample and ligated to the nucleic acids therein; the samples can then be pooled”).

6. The '810 and '828 Patents Anticipate the Asserted Claims or Render Them Obvious in Combination with Other Prior Art

(a) 1[pre]. A method of preparing nucleic acids for analysis, the method comprising:

732. Although I do not believe that the preamble is a limitation, to the extent that the preamble is a limitation, U.S. Patent Nos. 10,017,810 (“’810”)³²³ and 9,487,828 (“’828”)³²⁴, as well as Published Application No. US 2013/0303461³²⁵, Application No. 13/793,564 (the “’564 App.”) (collectively, the “’810/’828/’461/’564 Patents and Applications”) disclose the preamble.

733. The ’810/’828/’461/’564 Patents and Applications describe a method of preparing nucleic acid samples for sequencing. ’810 Col. 1:33-2:2; ’828 Col. 1:30-67; ’564 App. ¶¶ [0004]-[0006]. They describe methods to create sequencing libraries for various NGS sequencing platforms. *See, e.g.*, ’810 Col. 5:44-65; ’828 Col. 5:44-63 (describing Fig. 1); ’564 App. ¶ [0026]; *see also* ’810 Claims 1 and 16 (A method for preparing a nucleic acid for sequencing, the method comprising.”); ’828 Claims 1 and 40 (“a method for determining the nucleotide sequence

³²³ Ex. CCC

³²⁴ Ex. DDD

³²⁵ Ex. EEE

contiguous to a known target nucleotide sequence of 10 or more nucleotides, the method comprising”); ’564 App. Claim 1.

734. I have been informed that the Plaintiffs’ asserted patent claims rely on overboard interpretations and applications that are contrary to the Court’s Memorandum Order Regarding Claim Construction (’597 Patent) (D.I. 254). For the purposes of my review, I will rely on these interpretations and applications below.

(b) **1[a]. (a) contacting a first nucleic acid template comprising a sequence of a first strand of a double-stranded target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer;**

735. Taking into account the Court’s construction of the terms “double-stranded target nucleic acid,” “target nucleic acid,” “target-specific primer,” and “target-specific hybridization sequence,”³²⁶ the ’810/’828/’461/’564 Patents and Applications disclose this limitation. As set out above, the Court construed the term “double-stranded target nucleic acid” as “a target nucleic acid comprising a first strand and a second strand that are complementary to each other and hybridized to each other but that may have 3’ or 5’ overhangs or tails.”³²⁷ The Court construed the term “target nucleic acid” as “a nucleic acid molecule of interest (*e.g.*, a nucleic acid to be analyzed).”³²⁸ The Court construed the term “target-specific primer” to mean “a primer that has a level of complementarity between the primer and the target such that there exists an annealing temperature at which the primer will anneal to and mediate amplification of the target nucleic acid and will not anneal to or mediate amplification of non-target sequences present in a sample.”³²⁹ The Court

³²⁶ See D.I. 254, *Memorandum Order* at pp. 1-2.

³²⁷ See *id.* at p. 1.

³²⁸ See *id.*

³²⁹ See *id.* at pp. 1-2.

construed the term “target-specific hybridization sequence” to mean “a sequence of the target-specific primer that has sufficient complementarity with a sequence of the double-stranded target nucleic acid to enable hybridization between the target-specific primer and a sequence in/of the double-stranded target nucleic acid.”³³⁰

736. I have been informed that Plaintiffs allege that QIAGEN’s process of ligating a universal oligonucleotide tail adaptor to a target nucleotide sequence, and then using a first adaptor primer that anneals to that universal oligonucleotide tail adaptor and initiates a PCR reaction satisfies this element. The ‘810 and ‘828 specifications and claims make that same disclosure. *See e.g.*, ‘810 Col. 2:1-9; ‘828 Col. 2:1-9 (“(a) ligating a target nucleic acid comprising the known target nucleotide sequence with a universal oligonucleotide tail-adaptor; (b) amplifying a portion of the target nucleic acid and the amplification strand of the universal oligonucleotide tail adaptor with a first adaptor primer and a first target-specific primer...”); ‘564 App. ¶ [0007]. The specifics of the sequence of the universal oligonucleotide tail adaptor are described in the same column, and describe the universal oligonucleotide tail adaptor as having blocking and amplification strands that form a duplex and where the “amplification strand comprises nucleic acid sequences identical to a first and second sequencing primers.” *See* ‘810 Col. 2:13-26; ‘828 Col. 2:11-24; ‘564 App. ¶ [0007]. Further, the same section describes the first adaptor primer as “compris[ing] a nucleic acid sequence identical to a 5’ portion of the first sequencing primer.” ‘810 Col. 2:39-41; ‘828 Col. 2:37-39; ‘564 App. ¶ [0007]; *see also* ‘828 Col. 11:15-12:40 (identifying features of the “universal oligonucleotide tail adaptor”); ‘810 Col. 11:20-12:44; ‘564 App. ¶¶ [0064]-[0071].

737. The claims of the ‘810 and ‘828 patents have the same disclosures and descriptions. *See, e.g.*, ‘810 Claim 1(i) and (ii); Claim 16(i) and (ii); ‘828 Claim 1(i) and (ii) (“amplifying the

³³⁰ *See id.* at p. 2.

ligation product by polymerase chain reaction using a first adaptor primer that specifically anneals to the universal oligonucleotide tail-adaptor”); ’828 Claim 40 (same).

- (c) **1[b]. (b) contacting a second nucleic acid template comprising a sequence of a second strand that is complementary to the sequence of the first strand of the double-stranded target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers,**

738. The ’810/’828/’461/’564 Patents and Applications disclose this limitation, taking into account the Court’s construction of the phrases “a plurality of different primers” to mean “two or more different primers.”³³¹

739. I have been informed that Plaintiffs have interpreted this element to be met by QIAGEN’s use of gene specific primers that have a common 5’ tail. The same teachings exist in the ’810 and ’828 specification and claims. *See, e.g.*, ’810 Col. 2:8-11, ’828 Col. 2:6-9 (“amplifying a portion of the target nucleic acid and the amplification strand of the universal oligonucleotide tail adaptor with a first adaptor primer and a first target-specific primer...”); ’564 App. ¶ [0007].

740. The ’810 and ’828 specifications both describe there being a plurality of first gene specific primers with those plurality of primers having different hybridization sequences and a common sequence that is 5’ to the different hybridization sequences. *See, e.g.*, ’810 Col. 3:36-54; ’828 Col. 3:32-52 (“in some embodiments, the method can comprise contacting a single reaction mixture comprising the sample with a plurality of sets of first and second target specific primers. In some embodiments, the plurality of sets of first and second target-specific primers can specifically anneal to known target nucleotide sequences comprised by separate genes. In some

³³¹ *See* D.I. 254, *Memorandum Order* at p. 2.

embodiments, at least two sets of first and second target-specific primers can specifically anneal to different portions of a known target nucleotide sequences. In some embodiments, at least two sets of first and second target specific primers can specifically anneal to different portions of a single gene comprising a known target nucleotide sequence. . . . In some embodiments, the plurality of first target specific primers can comprise identical 5' tag sequence portions.”); ’564 App. ¶ [0013]; *see also* ’810 Col. 15:62-63 (“in some embodiments, the first target-specific primer can comprise a 5' tag sequence portion. In some embodiments, all first target specific primers present in a reaction can comprise identical 5' tag sequence portions.”); ’564 App. ¶ [0084].

741. The claims of the ’810 and ’828 also call for creating extension products formed by the plurality of first target specific primers that have the attributes discussed above. *See* ’810 Claim 1(ii), Claim 16 (ii); ’828 Claim 1 (ii); Claim 40 (ii); ’564 App. Claim 12.

(d) **1[c]. wherein the different hybridization sequences have different 3' ends, and wherein each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,**

742. The ’810/’828/’461/’564 Patents and Applications disclose this limitation, taking into account the parties’ agreement that the phrase “the same sequence” is to be interpreted as “the identical sequence.”³³²

743. The portions of the specification cited in the prior element state that there are a plurality of different primers that 1) anneal to sequences identifying separate genes, 2) different portions of a known target nucleotide sequence, 3) different portions of a single gene comprising known target nucleotide sequence. Each of these descriptions satisfy the requirement that the 3'

³³² *See* D.I. 254, *Memorandum Order* at p. 1.

ends of the plurality of different primers are different and not anneal to the same sequence of the double stranded target as Plaintiffs have been interpreting the claim.

- (e) **1[d]. wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and,**

744. Taking into account the Court's construction of the term "a sequence characteristic of the target-specific primer" to mean "a sequence from one of the strands of the single target nucleic acid recited in step (a),"³³³ the '810/'828/'461/'564 Patents and Applications disclose this limitation. *See, e.g.*, '810 Col. 2:11-12; '828 Col. 2:11-12; '564 App. ¶ [0007] ("(c)amplifying a portion of the amplicon resulting from step (b)"), '810 Col. 8:56-67; '828 Col. 8:53-64; '564 App. ¶ [0049] (defining amplicon as having sequence that is characteristic of the target and the primers used to perform the PCR reaction). *See also* Figs. 1 and 8 and specification describing them: '810 Col. 5:44-65, 6:22-23; '828 Col. 5:42-63; Col. 6:21-23; '564 App. ¶¶ [0026], [0032].

- (f) **1[e]. (c) subjecting the extension product to an amplification reaction comprising successive rounds of polymerase extension of i) a tail primer that comprises a 3' sequence that specifically anneals to the complement of the common sequence and that comprises a 5' tail sequence, and ii) a primer that specifically anneals to the complement of the target-specific hybridization sequence.**

745. I have been informed that Plaintiffs have alleged that the QIAGEN's products meet this limitation when its sample index primer anneals to common sequence on the first target specific primer and when the second adaptor primer anneals to sequence from the universal oligonucleotide tail adaptor.

746. I have been informed that Plaintiffs allege that the claims in the '810 patent cover the situation where the claimed second target specific primer anneals to any known sequence, including the common sequence from the first target specific primer. If this is the case, and there

³³³ *See* D.I. 254, *Memorandum Order* at p. 2.

is written description and enablement supporting that interpretation of the claims then the claims and specification of the '810 and '828 patents disclose this element. *See* '810 Col. 16:33-39; '828 Col. 16:29-35; '564 App. ¶¶ [0085] (describing the second target specific primer as comprising a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence.). *See also* '810 Claims 1 and 16 element (iii); '828 Claims 1 and 40 element (iii).

747. I have been informed that Plaintiffs have asserted that this particular element is met when the adaptor primer in QIAGEN's second round of PCR binds to sequence on or complementary to the universal oligonucleotide tail adaptor. This is disclosed in the specifications of the '810 and '828 patents as well as the '564 Application. *See, e.g.*, '810 Col. 2:11-13,41-44; '828 Col. 2:9-11, 39-44; '564 App. [0007] ("amplifying a portion of the amplicon resulting from step (b) with a second adaptor primer and a second target-specific primer. . . . and wherein the second adaptor primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer [in the universal oligonucleotide tail adaptor] and is nested with respect to the first adaptor primer."). *See also* '810 Claims 1 and 16 element (iii); '828 Claim 1 and 40 element (iii); '564 App. Claim 1 (c), and last "wherein" clause ("wherein the second adaptor primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first adaptor primer.").

(g) **3. The method of claim 1, wherein the first strand or the second strand of the double-stranded target nucleic acid is a deoxyribonucleic acid.**

748. The '810/'828/'461/'564 Patents and Applications disclose this limitation. *See* '810 Col. 3:10-11, 8:1-12; '828 Col. 3:8-9, 8:2-9; '564 App. ¶ [0010] ("in some embodiments, the sample can comprise genomic DNA"); see also Figs. 1 and 8 (identifying ds-cDNA or ds-genomic

I declare under penalty of perjury that the foregoing is true and correct.

Executed on October 7, 2020.

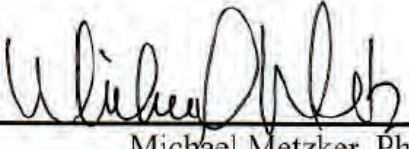

Michael Metzker, Ph.D.

EXHIBIT 4

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

ARCHERDX, LLC and THE GENERAL
HOSPITAL CORPORATION d/b/a
MASSACHUSETTS GENERAL HOSPITAL,

Plaintiffs,

v.

QIAGEN SCIENCES, LLC, QIAGEN LLC
f/k/a QIAGEN, INC., QIAGEN BEVERLY,
LLC f/k/a QIAGEN BEVERLY, INC.,
QIAGEN GAITHERSBURG, LLC f/k/a
QIAGEN GAITHERSBURG, INC., QIAGEN
GMBH, QIAGEN N.V., and
JONATHAN ARNOLD,

Defendants.

C.A. No. 18-1019-MN

REBUTTAL EXPERT REPORT OF NIAL L LENNON, PhD

Table of Contents

I.	Introduction.....	1
II.	QUALIFICATIONS	2
III.	SUMMARY OF OPINION	2
IV.	Legal Standard for Invalidity	2
A.	Burden of Proof.....	2
B.	Anticipation.....	3
C.	Obviousness	3
D.	Written Description	5
E.	Indefiniteness	6
V.	Claim Constructions.....	6
VI.	Materials Reviewed and Methodology	7
VII.	Summary of the Asserted Patents	7
VIII.	Claim Construction	8
IX.	Overview of the Alleged Prior Art References.....	13
A.	Alleged Prior Art Asserted Against the '810 Patent	14
1.	Amsterdam.....	14
2.	Siebert.....	15
3.	Markham.....	16
4.	Chenchik.....	22
5.	Kircher	25
B.	Alleged Prior Art Asserted Against the '597 Patent	30
1.	APAgene Manual and Fraiture.....	30
2.	McCarty	37
3.	Martin-Harris and Morley	40
4.	The '810 References.....	43
X.	Response to Invalidity Analysis: U.S. Patent No. 10,017,810	43
A.	Written Description	43
1.	Non-Nested Second Target-Specific Primer	43
2.	Second Target-Specific Primer Annealing to Sequence Other Than DNA to which Adaptors Are Ligated.....	48
3.	Sequence Identical to a Second Sequencing Primer.....	51
4.	Sequencing Primers in the Universal Oligonucleotide Tail Adaptor	51
5.	Second Adaptor Primer	52

B.	Indefiniteness	53
C.	Anticipation and Obviousness.....	57
1.	None of the Alleged Prior Art Discloses the Claimed Adaptor or Renders the Claimed Invention Obvious.....	57
2.	Amsterdam.....	60
a.	Claim 16[a]: Step (i)	60
(1)	Amsterdam Does Not Disclose an Amplification Strand Comprising Sequence Identical to a First and Second Sequencing Primers.....	61
(2)	The Proposed Combinations of Amsterdam with Markham or Kircher Do Not Render the Claimed Invention Obvious.	64
(3)	Amsterdam Does Not Disclose an Amplification Strand with a 3' T Overhang.....	72
b.	Claim 16[c]: Step (iii)	74
c.	Dependent Claims	75
3.	Siebert.....	76
a.	Claim 16[a]: Step (i)	76
(1)	Siebert Does Not Disclose an Amplification Strand Comprising Sequence Identical to a First and Second Sequencing Primers.....	76
(2)	Combining Siebert's Adaptors with the Illumina Sequences of Kircher Would Not Produce the Adaptor of the Claimed Method.	76
(3)	Combining Siebert's Adaptors with the Illumina Sequences of Amsterdam Would Not Produce the Adaptor of the Claimed Method.	81
(4)	Siebert Does Not Disclose a 3' T Overhang on an Amplification Strand	83
b.	Dependent Claims	84
4.	Markham.....	84
a.	Claim 16[a]: Step (i)	85
(1)	Markham Does Not Disclose an Amplification Strand Comprising Nucleic Acid Sequences Identical to a First and Second Sequencing Primers.	85
(2)	Combining Markham's Blocking Vectorette with Kircher's Adaptors Would Not Produce the Adaptor of the Claimed Method.	88
(3)	Combining Markham's Blocking Vectorette with Amsterdam's Adaptors and Wu's Two Rounds of LM-PCR Would Not Produce the Claimed Method.....	94
(4)	Markham Does Not Disclose an Amplification Strand with a 3' T Overhang.....	97
(5)	Dr. Metzker Fails to Identify A Plausible Motivation to Replace the Four-Base 5' Overhang on Markham's Blocking Vectorette with Bronner's TA Ligation Technique.....	100
(6)	Markham Does Not Disclose A Second Target-Specific Primer Having Sequence Identical To A Second Sequencing Primer.	105

b.	Dependent Claims.....	107
5.	Chenchik.....	108
a.	Claim 16[a]: Step (i)	108
(1)	Chenchik Does Not Disclose an Amplification Strand Comprising Nucleic Acid Sequences Identical to a First and Second Sequencing Primers.	108
(2)	Combining Chenchik's Adaptors with the Illumina Sequences of Kircher Would Not Produce the Adaptor of the Claimed Method.	110
(3)	Combining Chenchik's Adaptors with Amsterdam's Adaptors and Wu's LM-PCR Would Not Produce the Claimed Method.	112
(4)	Chenchik Does Not Disclose a 3' T Overhang on an Amplification Strand.....	113
b.	Claim 16[e]: Barcode	115
c.	Dependent Claims.....	117
XI.	Response to Invalidity Analysis: U.S. Patent No. 10,450,597	118
A.	Written Description	118
B.	Indefiniteness	121
C.	Anticipation and Obviousness.....	122
1.	The Degenerate Primer References Do Not Anticipate or Render Obvious the Claimed Invention of the '597 Patent	122
2.	Fraiture.....	127
a.	Claim 1[b]: A Plurality of Different Primers that Share a Common Sequence that Is 5' to Different Hybridization Sequences	127
b.	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends	130
c.	Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence	131
d.	Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence	136
e.	Dependent Claims.....	136
3.	McCarty	137
a.	Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence	137
b.	Dependent Claims.....	139
4.	Morley	140
a.	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends	140
b.	Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence	141

c.	Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence	142
d.	Dependent Claims.....	143
5.	Martin-Harris	143
a.	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends	143
b.	Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence	146
c.	Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence	150
d.	Dependent Claims.....	151
6.	APAgene Manual	152
a.	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends	152
b.	Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence	154
c.	Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence	155
d.	Dependent Claims.....	156
7.	'810 References	156
a.	Claim 1[e], Step (c)(i): A Tail Primer That Comprises a 3' Sequence that Specifically Anneals to the Complement of the Common Sequence and that Comprises a 5' Tail Sequence .	157
b.	Dependent Claims.....	162
XII.	Secondary Considerations of Non-Obviousness.....	162
a.	Commercial Success	163
b.	Industry Praise	165
c.	Long-Felt Need	167
d.	Skepticism.....	171
e.	Absence of Simultaneous Invention.....	172
f.	Conclusion	180

26. In my opinion, in the field of the patent which generally relates to next-generation DNA sequencing for use in determination of nucleotide sequences, a person of ordinary skill would typically have a Ph.D in Biology, Molecular Biology, Genetics, Biochemistry, Pharmacology, Biomedical Engineering or related fields, with 2-5 years of experience in next generation sequencing or a BSc in in Biology, Molecular Biology, Genetics, Biochemistry, Pharmacology, Biomedical engineering, or related fields and 5-6 years of experience in next generation sequencing.

VI. MATERIALS REVIEWED AND METHODOLOGY

27. Exhibit B to my Opening Report and Exhibit E to this report list documents and other data that I considered when forming my opinions. My opinions are also based in part upon my education, training, knowledge and experience.

28. In testifying, I may refer to some or all of the data or other documents and information identified herein, additional data or other information identified in discovery in this matter, as well as any materials relied upon by Qiagen or its experts, to support or summarize my opinions. In addition, I may supplement these materials with other materials, demonstrative exhibits, or charts to provide additional context, background, or information.

VII. SUMMARY OF THE ASSERTED PATENTS

29. The inventors of the '810 and '597 patents developed a methodology using single-primer extension (SPE) that allows users to analyze an adjacent unknown sequence next to a known sequence in a multiplex manner. Iafrate Depo. Tr. at 21:8-13; Le Depo. Tr. at 17:21-25.

30. The '810 patent is directed to methods of determining oligonucleotide sequences and enriching target sequences prior to sequencing the sequences. The claimed methods utilize a novel half-truncated adaptor, which includes a barcode portion, and which is ligated to nucleic

acids of interest. The adaptor-ligated nucleic acids are subjected to two rounds of amplification involving first and second target-specific primers and nested first and second adaptor primers.

31. The '597 patent is directed to methods for preparing and analyzing nucleic acids, methods for sequence analysis (e.g., using next-generating sequencing), and methods to enrich target nucleic acids prior to sequencing. The claimed methods of the '597 patent improve upon the '810 methods by utilizing a plurality of different primers having a common sequence that is 5' to different hybridization sequences, and which is utilized in successive rounds of PCR to amplify nucleic acid templates.

32. The patented SPE technique provides the following combination of benefits, which are detailed in Section XII of my Opening Report and which is incorporated by reference here:

- Detecting fusions with unknown fusion partners;
- Reducing PCR bias and errors;
- Small input amount;
- Barcoding;
- Efficiency and cost; and
- Detecting cancer with degraded input material.

VIII. CLAIM CONSTRUCTION

33. I understand that Qiagen has challenged claims 16-20 of the '810 patent against Qiagen.

34. Independent claim 16 of the '810 patent is reproduced below, with claim terms construed by the Court underlined:

same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers.” A POSA would understand that the negative limitation precludes two or more different primers of the plurality of different primers from annealing to the same sequence, no matter which sequence it may be.

317. Dr. Metzker’s analysis of the prior art also confirms that the terms “the first strand” and “the same sequence” do not render the ’597 claims indefinite. In his analysis of whether Fraiture, McCarty, Morley, Martin-Harris, the APAGene Manual and the ’810 References disclose “the first strand” and “the same sequence,” Dr. Metzker never raises the issue of which strand “the first strand” refers to or which sequence must be “the same sequence.” Nor does he indicate any difficulty understanding the scope of the ’597 claims in connection with analyzing whether the prior art anticipates or renders the claims obvious, much less assert that he lacked reasonable certainty as to how to interpret the claims with respect to the prior art.

C. Anticipation and Obviousness

318. Dr. Metzker contends that ’597 claims 1, 3, 4, 5, 10, 11, 13-27 and 29 are anticipated or rendered obvious by Fraiture, McCarty, Morley, Martin-Harris, the APAGene Manual, and the ’810/’828 patents, alone or in combination with each other and/or several additional references. Dr. Metzker fails to demonstrate that the prior art anticipates or renders obvious the challenged claims of the ’597 patent.

1. The Degenerate Primer References Do Not Anticipate or Render Obvious the Claimed Invention of the ’597 Patent

319. Dr. Metzker’s obviousness theory for five of the six primary references (i.e., Fraiture, McCarty, Morley, Martin-Harris and the APAGene Manual) involves reading the “plurality of different primers” onto primers having some degree of degeneracy, as summarized in the table below.

Reference	Primer	Hybridization Sequence	Permutations	Annealing Temps
Fraiture and APAGene Manual	DRT A	CCXXXGXXXXTAGT	262,144	42°C
	DRT B	CCXXXXXXTAGT	4,096	42°C
	DRT C	CCXXXXXXTCAT	4,096	42°C
	DRT D	CCXXXGXXXXCCTG	1,048,576	42°C
McCarty	Mu-TIR	CBCTCTTCKTCYATAATGGCAAT	12	54°C, 60°C
Martin-Harris and Morley	GW-A	NNNNNATTCG	1,024	35°C, 55°C
	GW-T	NNNNNTTTCG	1,024	35°C, 55°C
	GW-C	NNNNNCTTCG	1,024	35°C, 55°C
	GW-G	NNNNNGTTCG	1,024	35°C, 55°C

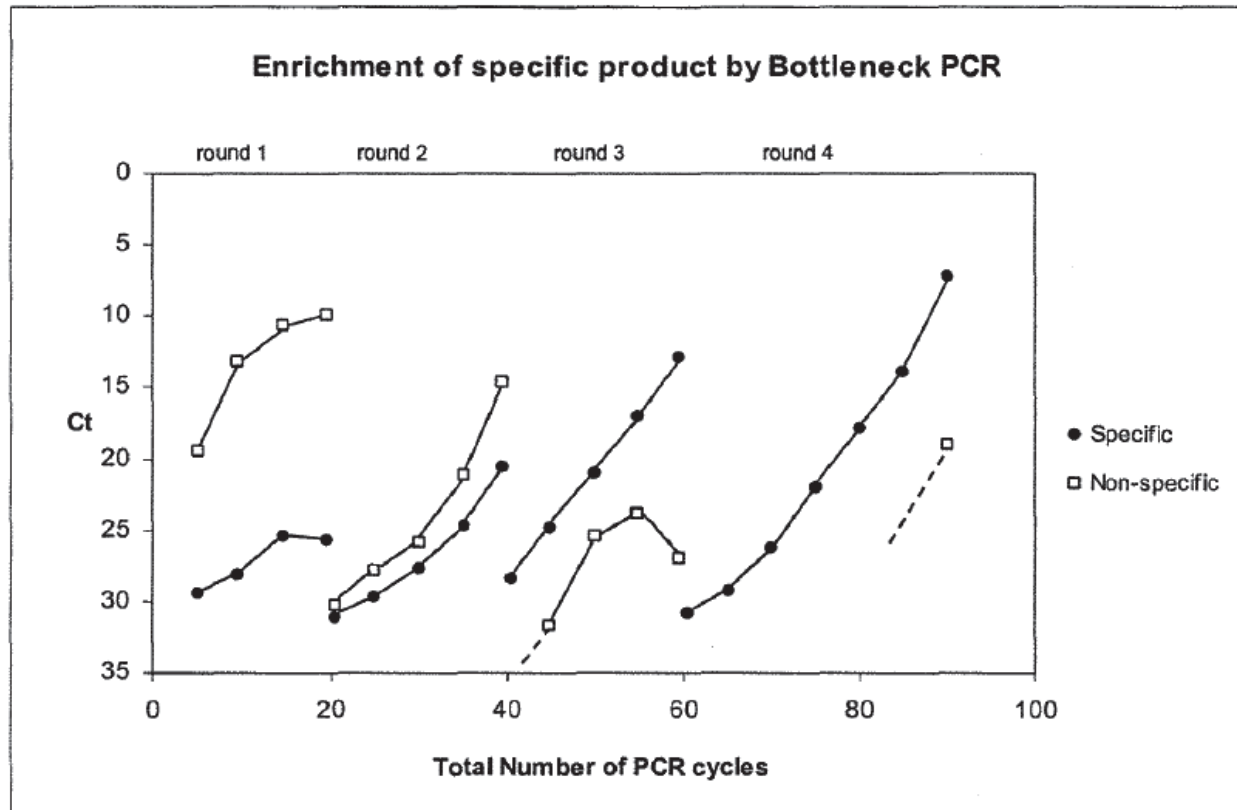
320. As Dr. Metzker explained in his 2001 article on PCR (Ex. H to Metzker Report), degenerate primers provide a strategy to obtain novel full-length cDNA sequences. *Id.*, p. 6. When attempting to sequence a region of interest, investigators sometimes have a good estimate of the likely sequence (e.g., by comparing orthologous genes to reveal conserved domains). *Id.* Degenerate primers can be used to create all possible permutations of regions or bases for which the sequence is unknown. Lowering the PCR annealing temperature increases the likelihood that a degenerate primer will anneal to a region of interest, even with less than 100% complementarity (i.e., even with some degree of mismatched bases).

321. Consistent with a typical degenerate primer strategy for amplifying a region of interest, the PCR conditions for the degenerate primer reactions in Fraiture, McCarty, Morley, Martin-Harris and the APAGene Manual utilize annealing temperatures below the 55°C to 68°C range that Dr. Metzker describes as “typical[]” (*id.*, p. 3), as detailed in the table above. In addition to low temperatures, the likelihood of mismatches is further increased by the number of degenerate bases and the degree to which each base is degenerate. The greater the degeneracy, the more complete the set of possible permutations, and therefore the greater the number of primers that differ by a single nucleotide.

322. Combining a large number of very similar primers with low temperature annealing conditions increases the likelihood that mismatches will occur. In a degenerate primer strategy, such mismatches are not necessarily an undesirable outcome for a researcher because mismatches increase the likelihood that any given region of interest will be amplified. Lowering the annealing temperature can increase the likelihood that at least some of the degenerate primers will amplify the region of interest even if none of the degenerate primers have 100% complementarity with the region of interest. This can occur, for example, if investigators design degenerate primers in which some of the “fixed” bases do not match a region of interest (e.g., due to an incorrect identification of a conserved region). While extension products may include an “incorrect” degenerate primer sequence (i.e., a sequence that differs from the region of interest), such inaccuracies need not affect sequencing because the degenerate primer sequence can be trimmed from the read, as described in McCarty (p. 3, Fig. 2 caption: “Validated reads are trimmed to remove the key and TIR sequences” where the 23-base TIR sequence includes the three partially degenerate bases B, K, and Y).

323. While mismatches and low annealing temperatures increase nonspecific products, specificity can be improved in subsequent rounds of PCR that “clean up” the degenerate primer products through a variety of techniques. For example, Fraiture, the APAGene Manual, Martin-Harris and Morley each use successive rounds of nested target-specific primers to reduce nonspecific products resulting from degenerate primers. In addition, Morley and Martin-Harris also utilize a “bottleneck” (or “inefficient priming”) technique to improve specificity over successive rounds. Morley Figure 5, shown below, illustrates how a first round of PCR with degenerate primers run at low temperatures (first 35°C, then 55°C) produces a greater abundance of non-specific products (illustrated as being detectable with fewer cycles [Ct]) than specific

product. The three subsequent rounds of “bottleneck” PCR, however, reduce the nonspecific products created by the degenerate primers until specific products predominate.



Morley, Figure 5.

324. The low-temperature annealing conditions of Dr. Metzker’s five prior art references and the use of degenerate primers (with some highly degenerate) would be expected to produce significant mismatches. A POSA would have expected at least some of the degenerate primers in these prior art references to anneal to the same sequence as some of the other degenerate primers.

325. Several definitions in the ’597 patent are relevant to evaluating whether the prior art satisfies the negative limitation requiring that “each primer of the plurality of different primers **does not anneal** to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,” as step (b) of ’597 claim 1 requires.

326. The term “anneal” is defined to mean “the formation of one or more complementary base pairs between two nucleic acids. In some embodiments, annealing involve[s] two complementary or substantially complementary nucleic acid strands hybridizing together.” 19:51-55. The term “complementary” is defined to mean “the ability of nucleotides to form hydrogen-bonded base pairs.” 37:56-57. In addition, the term “substantially complementary” is defined to mean “a nucleic acid molecule or portion thereof (e.g., a primer) having at least 90% complementarity over the entire length of the molecule or portion thereof with a second nucleotide sequence, e.g., 90% complementary, 95% complementary, 98% complementary, 99% complementary, or 100% complementary. 37:63-38:2.

327. Under these definitions, the scope of the negative limitation in step (b) of ’597 claim 1 broadly excludes different primers that anneal to the same sequence, including annealing where there is less than 100% complementarity between a primer and a template.

328. When discussing the limitation of step (b) of ’597 claim 1, Dr. Metzker testified that “if two primers share sufficient sequence homology, there is a good chance that they will anneal to the same sequence of a particular target under a particular set of conditions.” Dkt. No. 235, Declaration of Dr. Michael Metzker, ¶ 10. Dr. Metzker testified that “the ’597 Patent provides a way to estimate whether a particular primer is going to anneal to another polynucleotide, such as a sequence on a double-stranded target nucleic acid, under a number of conditions that take into account things like the length and sequence of the primer, its melting temperature, and salt concentration.” Id., ¶ 10 (’597 patent, 19:51-20:49). When analyzing step (b) of ’597 claim 1, however, Dr. Metzker did not consider any such information and made no attempt to estimate whether the degenerate primers in Fraiture, McCarty, Morley, Martin-Harris

and the APAGene Manual would or would not anneal to the same sequence on another polynucleotide.

329. Given the low-temperature PCR conditions of Qiagen's prior art references, the significant degree of similarity of the degenerate primers disclosed in the prior art references, the broad sweep of the negative limitation in step b of '597 claim 1, and Dr. Metzker's failure to even attempt to estimate whether any of the degenerate primers would anneal to the same sequence of target DNA, Dr. Metzker fails to establish by clear and convincing evidence that any of the degenerate primers in Fraiture, McCarty, Morley, Martin-Harris and the APAGene Manual ***do not anneal*** to the same sequence. Further supporting this conclusion is the analysis in the Expert Report of Dr. James McSwiggen affirmatively demonstrating that each of the prior art degenerate primers fails to disclose step (b) of '597 claim 1. Accordingly, Fraiture, McCarty, Morley, Martin-Harris and the APAGene Manual, alone or in combination, fail to anticipate or render obvious the '597 claims.

2. Fraiture

330. Dr. Metzker contends that '597 claims 1, 3, 4, 5, 10, 11, 13-27 and 29 are anticipated or rendered obvious by Fraiture in combination with the APAGene Manual and several additional references. Paragraphs 304-376.

a. Claim 1[b]: A Plurality of Different Primers that Share a Common Sequence that Is 5' to Different Hybridization Sequences

331. Step (b) of claim 1 requires "contacting a second nucleic acid template comprising a sequence of a second strand that is complementary to the sequence of the first strand of the double-stranded target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote

References were available as prior art to the '597 patent, they would not anticipate or render obvious the challenged claims of the '597 patent for the reasons detailed below.

a. Claim 1[e], Step (c)(i): A Tail Primer That Comprises a 3' Sequence that Specifically Anneals to the Complement of the Common Sequence and that Comprises a 5' Tail Sequence

417. Dr. Metzker contends that '597 claim 1, limitation (c)(i) ("a tail primer that comprises a 3' sequence that specifically anneals to the complement of the common sequence and that comprises a 5' tail sequence") is disclosed by the '810 References. Paragraph 746. The USPTO considered and rejected this argument. During prosecution the USPTO withdrew a rejection of '597 claim 1 based on the '828 patent after Plaintiffs added step (c), demonstrating that the limitations of step (c) make the '597 claims patentable over the '810 References.

418. During prosecution of the application that issued as the '597 patent, the USPTO examiner issued an office action dated January 31, 2018 finding that pending claim 1 was anticipated by the '828 patent. '597 File History, 1/31/2018 Office Action, p. 4. Claim 1, in the form it was pending at the time of that office action, is reproduced below.

1. A method of preparing nucleic acids for analysis, the method comprising:
 - (a) contacting a first nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer; and
 - (b) contacting a second nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers, wherein each primer of the plurality of different primers does not anneal to the same sequence of the target nucleic acid as any other primer of the plurality of different primers,
 wherein an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers.

'597 File History, 10/30/2017 Claims, p. 2.

419. Plaintiffs responded on July 31, 2018 by amending claim 1 to add step (c), as depicted below with additions indicated with underlining and deletions with strikethroughs:

1. A method of preparing nucleic acids for analysis, the method comprising:
 - (a) contacting a first nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer; ~~and~~
 - (b) contacting a second nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers, wherein each primer of the plurality of different primers does not anneal to the same sequence of the target nucleic acid as any other primer of the plurality of different primers,

wherein, following (a) and (b), an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers; and

(c) subjecting the extension product to an amplification reaction comprising successive rounds of polymerase extension of i) a tail primer that comprises a 3' sequence that specifically anneals to the complement of the common sequence and that comprises a 5' tail sequence, and ii) a primer that specifically anneals to the complement of the target-specific hybridization sequence.

'597 File History, 7/31/2018 Claims, p. 2.

420. In the next office action, dated September 7, 2018, the examiner responded: "The rejection of claims 1-18 and 20 under 35 USC 102(a)(1) as anticipated by [the '828 patent] has been withdrawn in view of the amendment." '597 File History, 9/7/2018 Office Action, p. 2.

421. In Paragraph 746, Dr. Metzker contends that step (c)(i) of '597 claim 1 is disclosed by the following three portions of the '810 References:

- '810 patent, 16:33-39; '828 patent, 16:29-35; '564 App. ¶ [0085]:

As used herein, the term "second target-specific primer" refers to a single-stranded oligonucleotide comprising a 3' portion comprising a nucleic acid sequence that can

specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (b), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer.

- '810 patent, claims 1 and 16 element (iii):

(iii) amplifying an amplification product of (ii) using a second target-specific primer that specifically anneals to the amplification product of (ii) and a second adaptor primer having a nucleotide sequence identical to a second portion of the amplification strand...

- '828 patent, claims 1 and 40 element (iii):

(iii) amplifying an amplification product of (ii) by polymerase chain reaction using a second adaptor primer and a second target-specific primer, wherein the second adaptor primer and the second target-specific primer are nested relative to the first adaptor primer and the first target-specific primer, respectively...

422. Dr. Metzker contends that if “there is written description and enablement supporting” the interpretation “that the claims in the '810 patent cover the situation where the claimed second target specific primer anneals to any known sequence, including the common sequence from the first target specific primer,” “then the claims and specification of the '810 and '828 patents disclose this element.” Paragraph 746.

423. I understand from counsel the following principles of patent law regarding the requirements for written description and anticipation that are relevant to analyzing Dr. Metzker's statement in Paragraph 746. For written description, I understand that:

- A specification need not describe every conceivable and possible future embodiment of their invention;
- A specification may provide adequate written description support of a broadly claimed invention without describing all species that the claim encompasses; and
- A specification need not describe unclaimed features of an accused product.

For anticipation, I understand that:

- The disclosure of a genus in the prior art is not necessarily a disclosure of every species that is a member of that genus; and
- To anticipate a later claim, a prior art reference must not only disclose all elements of the claim within the four corners of the document, but must also disclose those elements arranged as in the claim.

424. The above portions of the '810 References disclose a genus: a second target-specific primer (TSP2) that can specifically anneal to the amplification product of step (ii), including to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (ii). The above provisions do **not** disclose the particular species claimed in the '597 patent, i.e., they do not disclose amplifying an extension product of step (ii) with successive rounds of polymerase extension of a tail primer that comprises a 3' sequence that specifically anneals to the complement of the common sequence and that comprises a 5' tail sequence.

425. Step (iii) of '810 claim 16 claims a genus that includes “amplifying an amplification product of [step] (ii) using a second target specific primer that specifically anneals to the amplification product of [step] (ii).” The Court construed “second target-specific primer” to mean “a single-stranded oligonucleotide comprising a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (b), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer.” '810 Markman Order, p. 2.

426. Under the Court's construction, therefore, the element of claim 16 of step (iii) relevant to Dr. Metzker's statement in Paragraph 746 requires amplifying an amplification

product of step (ii) using a TSP2 that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (ii). The '810 claims do **not** require that the TSP2 anneal to any particular portion of the known target nucleotide sequence.

The '810 patent describes various species within this genus, including a TSP2:

- Annealing to separate genes (3:39-42);
- Annealing to different portions of a known target nucleotide sequence (3:42-45);
- Annealing to different portions of a single gene (3:45-49);
- Annealing to different exons of a gene (3:49-52);
- Annealing with at least 20, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, or 50 or more unique bases (16:53-60);
- Annealing at about 61°C to 72°C, 61°C to 69°C, 63°C to 69°C, 63°C to 67°C, 64°C to 66°C, or specifically at about 65°C in a PCR buffer (18:48-56).

427. As the above list illustrates, the '810 patent discloses a broad variety of species within the genus of ways that a TSP2 can anneal to an amplicon of step (ii). However, the '810 patent does not disclose the particular combination of elements in step (c) of '597 claim 1, which requires (1) amplifying an extension product with (2) a tail primer that includes (3) a 5' tail sequence and (4) a particular 3' sequence (5) that specifically anneals to a portion of the extension product comprising the complement of a common sequence, (6) which common sequence was introduced into the extension product via the 5' end of a plurality of different primers.

428. Dr. Metzker argues that "the '810 and '828 patents disclose this element," i.e., step (c) of '597 claim 1 but provides no supporting evidence beyond citing the definition of TSP2 (Paragraph 746), which simply recites the genus disclosed in the '810 patent. Dr. Metzker

provides no analysis whether step (c) of '597 claim 1 would have been obvious in view of the '810 References. Dr. Metzker acknowledges that the obviousness determination requires evaluating differences between the claims and the prior art (Paragraph 39), but he fails to evaluate the differences between the claimed species of step (c) of '597 claim 1 and the disclosed genus of the '810 References. Nor does Dr. Metzker contend that a POSA would have been motivated to combine the '810 References with any prior art or with the knowledge of a POSA.

429. Accordingly, the '810 References do not disclose amplifying an extension product of step (ii) with successive rounds of polymerase extension of a tail primer that comprises a 3' sequence that specifically anneals to the complement of the common sequence and that comprises a 5' tail sequence.

430. For the foregoing reasons, the Dr. Metzker fails to establish by clear and convincing evidence that the '810 References anticipate or render obvious '597 claim 1.

b. Dependent Claims

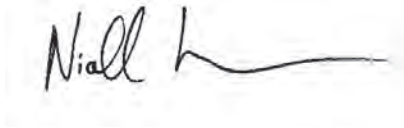
431. Dependent claims 3-5, 10, 11, 13-27, and 29 of the '597 patent include each of the elements of Claim 1. The fatal flaws in Dr. Metzker's anticipation and obviousness analysis regarding claim 1 identified above are fatal to his analysis of the dependent claims for the same reasons. Accordingly, Dr. Metzker fails to establish that dependent claims 3-5, 10, 11, 13-27, and 29 are anticipated or obvious.

432. For the foregoing reasons, Dr. Metzker fails to demonstrate that the '810 References, alone or in combination with the prior art references identified in Dr. Metzker's report, anticipates or renders obvious the challenged claims of the '597 patent.

XII. SECONDARY CONSIDERATIONS OF NON-OBVIOUSNESS

433. The inventors of the '810 and '597 patents developed a methodology using single-primer extension (SPE) that allows users to analyze an adjacent unknown sequence next to a

I declare under penalty of perjury that the foregoing is true and correct.

A handwritten signature in black ink. The name "Niall" is written in a cursive script, followed by a stylized "H" that extends into a long horizontal line.

Dated: 11-20-2020

EXHIBIT 5

Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS)

C.R.Newton, A.Graham, L.E.Heptinstall, S.J.Powell, C.Summers, N.Kalsheker¹, J.C.Smith and A.F.Markham

ICI Diagnostics, Gadbrook Park, Northwich, Cheshire CW9 7RA, ICI Cellmark Diagnostics, Blacklands Way, Abingdon Business Park, Abingdon, Oxon OX14 1DY, ¹University of Wales College of Medicine, Department of Medical Biochemistry, Newport Road, Cardiff CF2 1SZ, UK and ICI Cellmark Diagnostics, 20271 Golden Rod Lane, Germantown, MD 20874, USA

Received February 23, 1989; Accepted March 2, 1989

ABSTRACT

We have improved the "polymerase chain reaction" (PCR) to permit rapid analysis of any known mutation in genomic DNA. We demonstrate a system, ARMS (Amplification Refractory Mutation System), that allows genotyping solely by inspection of reaction mixtures after agarose gel electrophoresis. The system is simple, reliable and non-isotopic. It will clearly distinguish heterozygotes at a locus from homozygotes for either allele. The system requires neither restriction enzyme digestion, allele-specific oligonucleotides as conventionally applied, nor the sequence analysis of PCR products. The basis of the invention is that unexpectedly, oligonucleotides with a mismatched 3'-residue will not function as primers in the PCR under appropriate conditions. We have analysed DNA from patients with α -antitrypsin (AAT) deficiency, from carriers of the disease and from normal individuals. Our findings are in complete agreement with allele assignments derived by direct sequencing of PCR products.

INTRODUCTION

The analysis of nucleic acid sequence is central to biology. Determination of variation in DNA sequence between individuals underpins molecular genetics. Such analysis is routinely performed by examination of restriction fragment length polymorphism (RFLP) using the Southern blotting technique (1,2,3). This approach has proved enormously useful, generating a massive literature, despite the fact that it is relatively slow and only allows for the examination of the limited number of polymorphic base changes which either create or destroy a restriction endonuclease recognition site. Without doubt any method which enabled all polymorphic base changes in the genome to be examined in a facile manner would be invaluable to the molecular genetics community.

Nucleic Acids Research

PCR (4) has greatly facilitated the analysis of genomic DNA. It allows diagnosis of genetic diseases when combined with one of a variety of other techniques (5,6,7,8,9,10,11,12). We and others have reported the use of PCR and direct sequencing for diagnosis of inherited diseases (5,6,7,13). Allele-specific oligonucleotides (ASOs), (14,15) either radio-labelled (8) or non-isotopically tagged (9) have been applied to disease diagnosis in the conventional manner by probing dot blots of PCR products. Occasionally a point mutation giving rise to a specific phenotype may create or destroy a restriction enzyme recognition site (2). In such instances PCR products may (or may not) be cleaved when treated with the restriction enzyme. The presence or absence of the restriction site can be used to perform diagnoses as recently demonstrated for sickle cell anaemia (10). Similarly a polymorphic restriction site may be in linkage with an uncharacterised mutation allowing diagnoses to be performed in informative families by analysis of the amplified restriction site polymorphism (11,16,17).

We demonstrate here a general technique which allows the scrutiny of any point mutation polymorphism. The technique requires that the terminal 3'-nucleotide only of a PCR primer be allele specific. Thus the primer is synthesised in two forms. The 'normal' form is refractory to PCR on 'mutant' template DNA and the 'mutant' form is refractory to PCR on 'normal' DNA. In some instances a single 3'-mismatched base does allow amplification to proceed. We have shown that introducing additional deliberate mismatches near the 3' end of appropriate primers ameliorates this problem.

Molecular characterisation of the genes associated with the more common inherited disorders is constantly providing new information about the underlying, disease-associated mutations. Indeed recent sequencing of mutant β -globin genes has only rarely resulted in the discovery of novel alleles (18). This implies that at the β -globin locus characterisation of the molecular pathology is nearing completion (13). Diseases such as cystic fibrosis, as yet uncharacterised at the gene level, may have several RFLPs in linkage disequilibrium with the affected phenotype (19). Such RFLPs are useful for haplotype analysis and

risk assessment of carrier status particularly where there is a family history of the disease (20). Furthermore, flanking sequences of some such RFLPs have been determined allowing PCR followed by restriction analysis for haplotype identification (16,17). Some concern has been expressed as to the reliability and reproducibility of such assays in the absence of rigorous and appropriate internal controls. In theory ARMS would allow rapid haplotype analysis in such situations, providing sufficient genomic sequence is known around the polymorphic restriction site.

The feasibility of our ARMS was demonstrated by the amplification of exon III and part of intron III in the human AAT gene (figure 1). Direct application of ARMS to the clinically significant S and Z alleles of AAT (21) was performed and the diagnoses were in agreement with the results of sequence analysis of the PCR products (5).

MATERIALS AND METHODS

DNA preparation

Genomic DNA was isolated from peripheral blood cells as described previously (5).

Oligonucleotide amplification, amplification refractory and sequencing primers

The common primers 1,2,5 and 6 (figure 1) were those described previously (5). Their respective sequences were d(CCCACCTTCCCCTCTCTCCAGGCAAATGGG), d(GGGCCTCAGTCCCAACATGGCTAAGAGGTG), d(TGTCCACGTGAGCCTTGCTCGAGGCCTGGG) and d(GAGACTTGGTATTTGTTCATCATTAAG). Primer 2a and the 3,4,7 and 8 series of primers (figure 2) as well as the primers for the internal control, a 510 base pair fragment from the unusually long exon 26 of the human apolipoprotein B gene (22) were prepared as described (5) and were used without further purification. The sequencing primers for initial allele characterisation were those described earlier (5).

Allele characterisation by PCR and direct sequencing

Mutant and normal alleles of the AAT S and Z loci were confirmed by PCR amplification either as described (5), or as follows; target sequences were amplified in a 100µl reaction volume

Nucleic Acids Research

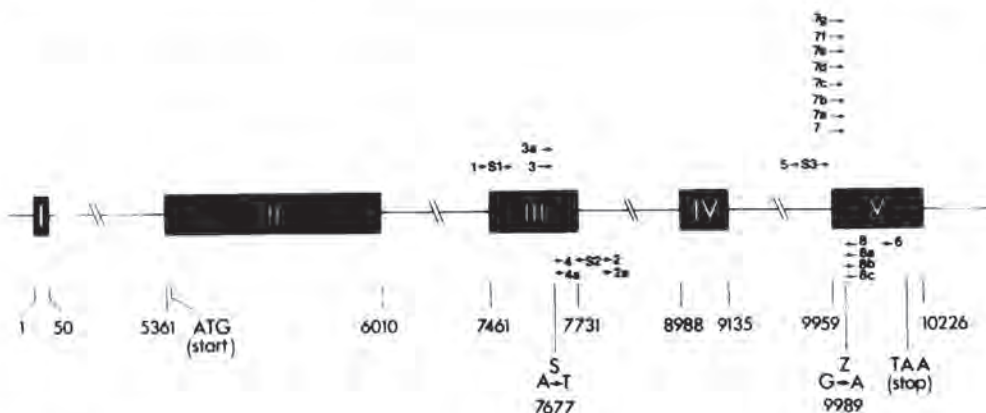


Figure 1

The human alpha-1-antitrypsin gene. Coordinates are as described by Long *et al.* (24). Position 1 is the proposed transcription start site. The solid boxes represent the five exons, the S and Z loci are shown, as are the respective mutations responsible for the S and Z phenotypes. The arrows below and above the gene represent the various primers used. Primers prefixed by S are those used in direct sequencing of PCR products to confirm genotypes prior to ARMS analyses. The remaining primers are those used to demonstrate the feasibility of the ARMS concept and those used in ARMS analyses, these primers are shown in detail in figure 2.

containing approximately 1µg genomic DNA, deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP) and thymidine triphosphate (TTP), each 1.5mM, 67mM Tris-HCl (pH8.8), 16.6mM ammonium sulphate, 6.7 mM magnesium chloride, 10mM 2-mercaptoethanol, 6.7µM EDTA and 1µM each appropriate amplification primer. Samples were heated at 100°C for 5 minutes to denature the DNA. Two units of *Thermus aquaticus* (Taq) DNA polymerase (23) (Anglian Biotechnology) were added to each sample. Samples were overlaid with light mineral oil (Sigma, 50µl) then heated at 60°C for 4 minutes for the first round of DNA synthesis. (See Discussion). Subsequent cycles consisted of a two minute denaturation step at 92°C and a combined primer annealing and DNA synthesis step at 60°C for 4 minutes. 33 cycles were performed and the DNA synthesis step of the final cycle was extended to 20 minutes. Direct sequencing of the PCR products was as described previously (5).

ARMS analysis of genomic DNA

The feasibility of the ARMS concept was demonstrated using duplicate samples of genomic DNA from one normal individual. AAT exon III primers 1 and 2 were used with one sample; primers 1 and 2a (3'C/T mismatch) were used with the other sample. AAT exon V primers 5 and 6 were present in both samples serving as an internal control. All primers are shown in figure 1. PCR reactions were performed and examined by agarose gel electrophoresis (3% Nu-sieve) as previously described (5).

In applying ARMS to subsequent mutation analysis, primers 'Control 1' d(CTCTGGGAGCACAGTACGAAAAACCACTT) and 'Control 2' d(AA TGAATTTATCAGCCAAAACCTTTACAGG) were included in all reactions and served to provide an internal control PCR product. The control primers amplify a 510 base pair product within exon 26 of the human apolipoprotein B gene (22).

Genomic DNAs of characterised AAT genotypes MM,MS,MZ and ZZ were subjected to PCR so as to amplify the internal control fragment. In separate pairs of reactions each DNA was either coamplified with the appropriate 'normal' or 'mutant' primer paired with a common primer for the respective AAT locus. These primers are shown in figure 2.

The reactions for the ARMS analyses were performed in a volume of 100µl containing approximately 1µg genomic DNA. dATP, dCTP, dGTP and TTP were each 1.5mM in 67mM Tris-HCl (pH8.8), 16.6mM ammonium sulphate, 6.7mM magnesium chloride, 10mM 2-mercaptoethanol, 6.7µM EDTA and 1µM each appropriate amplification primer. Samples were heated at 100°C for 5 minutes to denature the DNA. Two units Taq DNA polymerase (Anglian Biotechnology) was added to each sample. Samples were overlaid with light mineral oil (Sigma, 50µl) then heated at 60°C for 4 minutes for the first round of DNA synthesis. Subsequent rounds of amplification comprised two minutes denaturation at 92°C followed by combined primer annealing and DNA synthesis at 60°C for four minutes. 33 cycles were performed in this way with the final synthesis step extended to 20 minutes. 18µl from each reaction was combined with 2µl of 50% glycerol 0.2% bromophenol blue in 1X TBE then electrophoresed on 1.4% agarose gels in 1X TBE containing 0.5µg/ml ethidium bromide.

Nucleic Acids Research

```

5' ATTCCTCAACCTGAGGGTGACCAAGAAGCTGCCACACCTCTTAGCCATGTTGGGACTGAGGCCCATCAGGACTGGC 3'
3' TAAGGGTTGGACTCCCACTGGTTCCTCGACGGGTGTGGAGAATCGGTACAACCCTGACTCCGGGTAGTCCTGACCG 5'
|
7735 Primer 2 GTGGAGAATCGGTACAACCCTGACTCCGGG 5'
Primer 2a TTGGAGAATCGGTACAACCCTGACTCCGGG 5' 7811

5' GCCTGATGAGGGGAACTACAGCACCTGGT Primer 3a
5' GCCTGATGAGGGGAACTACAGCACCTGGA Primer 3
5' CTCCTGCCTGATGAGGGGAACTACAGCACCTGGAATAAGTGAACCAACGATATCATACCAAGTTCCTGGAAA 3'
3' GAAGGACGGACTACTCCCTTTGATGTCGTGGACCTTTACTTGAGTGGGTGCTATAGTAGTGGTTCAAGGACCTTT 5'
|
7642 Primer 4 TTTTACTTGAGTGGGTGCTATAGTAGTGGT 5'
Primer 4a ATTTACTTGAGTGGGTGCTATAGTAGTGGT 5' 7718

5' CCGTGCATAAGGCTGTGCTGACCTCGACA Primer 7g
5' CCGTGCATAAGGCTGTGCTGACCTCGACG Primer 7f
5' CCGTGCATAAGGCTGTGCTGACCATAGACA Primer 7e
5' CCGTGCATAAGGCTGTGCTGACCATAGACG Primer 7d
5' CCGTGCATAAGGCTGTGCTGACCATCGCCA Primer 7c
5' CCGTGCATAAGGCTGTGCTGACCATCGCG Primer 7b
5' CCGTGCATAAGGCTGTGCTGACCATCGACA Primer 7a
5' CCGTGCATAAGGCTGTGCTGACCATCGACG primer 7
5' TCCAGGCCGTGCATAAGGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTAGAGGCC 3'
3' AGGTCCGGCACGTATTCCGACACGACTGGTAGCTGcTCTTTCCCTGACTTCGACGACCCCGGTACAAAAATCTCCGG 5'
|
9954 Primer B CTCTTCCCTGACTTCGACGACCCCGGTAC 5'
Primer 8a TTCTTCCCTGACTTCGACGACCCCGGTAC 5' 10030
Primer 8b CTCATTCCTGACTTCGACGACCCCGGTAC 5'
Primer 8c TTCATTCCTGACTTCGACGACCCCGGTAC 5'

```

Figure 2

ARMS primers. The top panel shows the primers used to test the ARMS concept. Primer 2 is complementary to the coding strand of the AAT gene. Primer 2a shows the 3'-OH mismatched T residue. Primers 2 and 2a are used in conjunction with primer 1 (figure 1). The centre panel shows the ARMS primers employed at the AAT S locus. The lower case A/T base pair is the AAT S locus and the depicted sequence is the normal sequence. The AAT S variant DNA has a T/A base pair at this position. Primers 3 and 4 correspond to 'normal' sequence, primers 3a and 4a correspond to 'mutant' sequence. The lower panel shows the the ARMS primers employed at the AAT Z locus. The lower case G/C base pair is the Z locus and the normal sequence is shown. The AAT Z variant DNA has an A/T base pair at this position. Primers 7 and 8 correspond to the 'normal' sequence, primers 7a and 8a correspond to the 'mutant' sequence. Primers 7, 8, 7a and 8a have not been destabilised. The remaining primers in the 7 and 8 series are destabilised and the deliberately introduced mismatches are underlined. Primers 7b, 7d, 7f and 8b correspond to 'normal' sequence (discounting the deliberate mismatches) likewise primers 7c, 7e, 7g and 8c correspond to 'mutant' sequence, again discounting the introduced mismatches. The position numbers are as described by Long et al. (24).

RESULTS

ARMS primers

Figure 1 shows the ARMS primers in relation to the human AAT gene. Figure 2 shows each ARMS primer sequence in detail with respect to the gene. The position numbers are measured from the proposed transcription start site of the AAT gene (24). Discounting the variable 3' nucleotides and deliberately introduced mismatches, the 2,3,7 and 8 series primers are 59% GC 30mers and the 4 series primers are 41% GC 30mers. The common primers 1,2,5 and 6 are a 63% GC 30 mer, 60% GC 30mer, 67% GC 30mer and a 31% GC 29mer respectively.

Feasibility of the ARMS concept

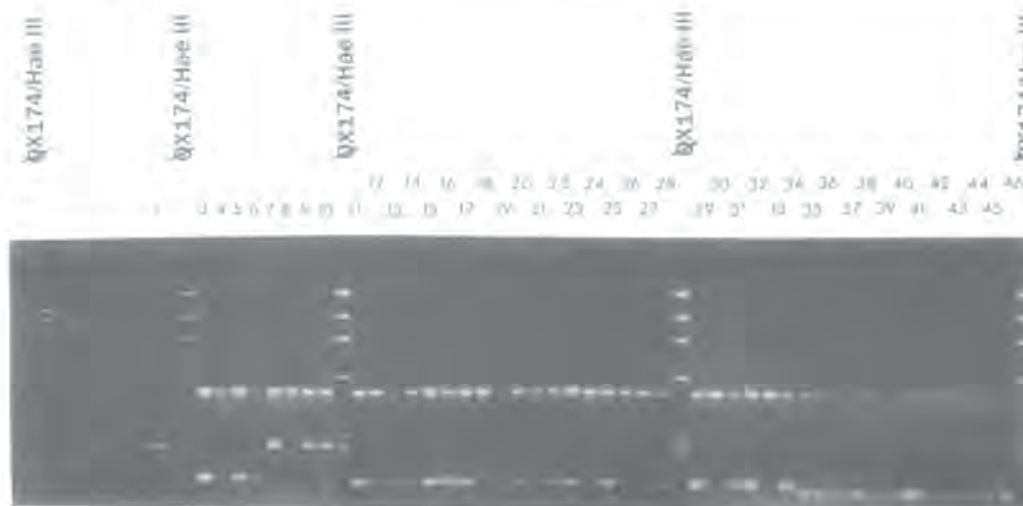
We have shown previously that the AAT gene regions bounded by primers 1 and 2 and by primers 5 and 6 (figures 1 and 2) can be coamplified without affecting the efficiency of amplification of either target performed in isolation (5). We chose to introduce a 3' terminal base change into primer 2. Specifically the 3' dG residue was replaced by T to provide primer 2a. This substitution generates a template/primer C/T mismatch. When primers 1,2,5 and 6 are combined in a PCR, both the 360 bp product bounded by primers 1 and 2, and the 220 bp product bounded by primers 5 and 6 are observed (figure 3, lane 1). Substitution of primer 2 by primer 2a however blocks amplification of the 360 bp product while the internal control 220 bp product is still generated (figure 3, lane 2). This result is attributable to the lack of a 3' exonucleolytic proofreading activity of Taq DNA polymerase (23) in agreement with the observations of Tindall and Kunkel (25).

ARMS analysis of the AAT S locus

Genomic DNAs, either homozygous normal with respect to the AAT gene S allele or heterozygous S were each amplified as described in Materials and Methods. Each DNA was separately amplified using primers 2 and 3 and primers 2 and 3a. Primer 3 corresponds to the normal sequence at the S locus and primer 3a corresponds to the S variant sequence. In all reactions the internal control primers were also included.

On the normal DNA substrate, product was derived only from the internal control primers and primers 2 and 3. No product was

Nucleic Acids Research

**Figure 3**

Agarose gels showing the feasibility of the ARMS concept (lanes 1 and 2) and ARMS analyses at the AAT S locus (lanes 3 to 10) and AAT Z locus (lanes 11 to 46). Specific reactions are as described in the text (results section).

observed when primer 3a replaced primer 3. When heterozygous S DNA replaced homozygous normal DNA the expected 152 bp product was generated when either primer 3 or 3a was included in the reaction (figure 3, lanes 3 to 6). Primer 3 generates an A/A mismatch with S variant DNA and primer 3a generates a T/T mismatch with normal DNA. When the ARMS detection primers were designed for the opposing strand at the S locus (primers 4, normal and 4a, S variant) and used for amplification with a common primer 1, similar results were obtained. The 510 bp internal control was generated but the 267 bp product was observed only when the normal primer was applied to normal DNA or the normal or S variant primer was applied to heterozygous S DNA. The 267 bp product was not generated when the S variant primer was applied to normal DNA (figure 3, lanes 7 to 10). Primer 4 generates a T/T mismatch with S variant DNA and primer 4a generates an A/A mismatch with normal DNA.

ARMS analysis of the AAT Z locus

In analogous experiments to the ARMS analyses of the AAT S locus we amplified genomic DNAs characterised as normal, heterozygous and homozygous at the AAT Z locus. All reactions contained the

internal control primers 1 and 2. Initial experiments contained primer 7 (normal) or 7a (mutant) for amplification with common primer 6 to yield a 150 bp product (figures 1 and 2). Alternative experiments targetting the opposing strand and employing primers 8 (normal) or 8a (mutant) for amplification with common primer 5 (figures 1 and 2) would give products of 129 bp. Figure 3, lanes 11-16 and lanes 35-40 shows the products generated by the respective use of primer 7 with normal DNA, primer 7a with normal DNA, primer 7 with heterozygous DNA, primer 7a with heterozygous DNA, primer 7 with homozygous Z (ZZ) DNA, primer 7a with ZZ DNA, primer 8 with normal DNA, primer 8a with normal DNA, primer 8 with heterozygous DNA, primer 8a with heterozygous DNA, primer 8 with ZZ DNA and primer 8a with ZZ DNA. In contrast to the ARMS data for the AAT S locus, the corresponding results for the AAT Z locus show reduced specificity in that products were evident using either normal primer with ZZ DNA and either mutant primer with normal DNA. Primer 7 with ZZ DNA generates a primer/template G/T mismatch. Conversely primer 7a with normal DNA generates an A/C mismatch. Primer 8 with ZZ DNA generates a C/A mismatch and primer 8a with normal DNA generates a T/G mismatch.

In an attempt to increase the specificity of the ARMS primers we chose to deliberately introduce an additional mismatch near their 3'-ends. When primers 7f and 7g which have a deliberate C/T mismatch seven bases from the 3'-end were introduced to replace primers 7 and 7a (figure 2) specificity was improved. Figure 3, lanes 17-22 shows the products of these reactions. In particular, lane 18 shows the virtual absence of the 150 bp product when the mutant primer (7g) is applied to normal DNA. Unfortunately the normal primer (7f) still generates a small amount of product with ZZ DNA (figure 3, lane 21), but much reduced with respect to the yield with the mutant primer on ZZ DNA with equivalent internal control products. When primers 7d and 7e which have a deliberate A/G mismatch five bases from their 3'-ends were introduced to replace primers 7 and 7a similar results were obtained (figure 3 lanes 23 to 28) to those with primers 7f and 7g. When primers 7b and 7c replaced primers 7 and 7a in the system the desired specificity was observed. Primers 7b and 7c have a deliberate C/T mismatch three bases from their 3'-ends. Specifically only primer

Nucleic Acids Research

7b generated a 150 bp product with normal DNA (fig. 3 lane 29). The 'mutant' primer 7c failed to do so (fig. 3 lane 30). Both primers 7b and 7c generated product from heterozygous DNA (fig. 3 lanes 31 and 32). Primer 7b failed to generate the 150 bp product with ZZ DNA whereas the 'mutant' primer 7c did generate the 150 bp product (fig. 3 lanes 33 and 34). Similar exchange of primers 8 and 8a for primers 8b and 8c (which have an A/A mismatch four bases from their 3'-ends) showed marginal increased specificity (fig. 3 lanes 41 to 46).

DISCUSSION

Interest is increasingly being focused on the mutations in the human genome which produce disease states. The number of such mutations characterised at the DNA sequence level is increasing rapidly and this has been substantially aided by the PCR/direct sequencing approach for the analysis of genomic DNA. We previously reported that PCR followed by direct sequencing was absolutely specific for the diagnosis of AAT deficiency (5). In this communication we present ARMS, a system allowing the direct analysis of any locus of interest and thus generally applicable to any inherited disease provided sufficient sequence data is available. ARMS is simple, rapid and reliable providing the capability for accurate pre- and postnatal diagnosis and a means for heterozygote detection. ARMS is still of benefit even if disease-associated mutations, as yet uncharacterised, are linked to characterised polymorphisms. In such instances the technique will allow detailed haplotype analyses to be performed with a minimal quantity of DNA. Accurate prenatal diagnoses are achievable in a few hours if maternal contamination of the foetal material is avoided. An important practical consideration with this approach (as with other PCR-based strategies) is that it is unnecessary to prepare high quality DNA suitable for restriction enzyme digestion.

A prerequisite of ARMS is the absence of a 3'-exonucleolytic proofreading activity associated with the DNA polymerase employed. The lack of such an exonuclease associated with Taq DNA polymerase has been confirmed here by the successful application of ARMS and independently (23,25). Another requirement in the

application of ARMS is that 3'-OH terminal mismatched primers are refractory to extension by the chosen DNA polymerase. This was not apparent from the work of Tindall and Kunkel (25), since their exonuclease assay required, and did generate, polymerase products from C/A mismatched primer/template complexes. Taq polymerase refractory mismatches have been demonstrated in this work for some mismatched primers. In instances where the mismatch is not refractory to extension (as demonstrated with primer/template G/T, A/C, C/A and T/G mismatches at the AAT Z locus) further deliberate mismatches to destabilise the primer/template complexes render the primers increasingly refractory as the additional mismatch is moved progressively closer to the 3' end.

Empirically, the degree of specificity observed with mismatched primers (and thus the requirement for additional destabilisation), correlates with the mismatch type. C/T, A/A and T/T mismatches (which are all either purine/purine or pyrimidine/pyrimidine mismatches) are considerably more refractory to extension by Taq polymerase than G/T, T/G, A/C or C/A mispairs (all purine/pyrimidine mismatches). We have not yet optimised the position for introduction of the deliberate mismatches, nor the type of mismatch, neither have we examined the effect of deliberate base-pair insertions, deletions or modifications which may also be expected to appropriately destabilise non-refractory primers.

It is likely that any of these approaches to deliberately destabilise the ARMS primers and hence improve specificity may be enhanced by reducing dNTP, magnesium and primer concentrations or simply increasing the amplification annealing/extension temperature. Conversely an increase in concentration of these reagents or a decreased amplification annealing temperature might be expected to have an adverse effect on specificities of primers which previously generated no unwanted product.

We have deliberately chosen relatively unforgiving amplification conditions in this series of experiments so as to challenge the basic concept fully. Removal of tubes from the heating block to facilitate the addition of enzyme after initial DNA denaturation at 100°C was performed in the experiments described herein.

Nucleic Acids Research

Undoubtedly this allows cooling of reaction mixtures and is difficult to control precisely. Products were generated at the AAT Z locus with mismatched, non-destabilised primers and primers with additional mismatches 7, 5 and 4 bases from their 3'-termini. It is conceivable that the generation of these products will be avoided, if, after heat denaturation of the genomic DNA in the presence of primers the reactions are not allowed to briefly cool during enzyme addition. These products may result from a proportion of template molecules being primed at a less stringent lower temperature than the routine extension temperature. Any extension products so derived would then be correctly paired with the ARMS primer in subsequent cycles, and so generate the observed unwanted products.

Destabilisation of the ARMS primers where necessary such that anomalous products are not generated has been one of our approaches to the development of this technique at this early semi-manual stage. Obviously, further refinement is possible by optimising such variables as magnesium, dNTP or Taq polymerase concentrations and the precise temperature throughout the ARMS cycles. Careful control of the later variable in particular should be achievable with fully automated ARMS instrumentation. Indeed, addition of Taq polymerase to ARMS reaction mixtures at 60°C and ensuring that the reaction temperature never falls below this, may significantly increase reaction specificity and avoid generation of products on mismatched templates.

The AAT Z mutation is caused by a G to A transition immediately preceded by a C residue. The AAT S phenotype results from an A to T mutation. Analysis of single base mutations within coding regions causing human genetic disease shows that 35% of such mutations have occurred within CpG dinucleotides and that over 90% of these were either C to T or G to A transitions (26). Since such mutations would generate the same primer/template mismatches as at the AAT Z locus, it is expected that destabilisation of ARMS primers will be required for at least 30% of ARMS potential applications.

We have chosen to use large (30mer) primers in this assay since this allows the use of high annealing temperatures to improve specificity and reduces the chance of mispriming elsewhere on

genomic DNA. An alternative approach might be to use shorter primers which span a point mutation such that discrimination between 'normal' and 'mutant' loci is achieved by hybridisation of the primers in an allele-specific manner under appropriately stringent conditions. This type of analysis, also of the AAT Z locus, has been performed by Dermer and Johnson (27). It is important to note however that in this analysis the specificity of the primers was not absolute, the absence of internal controls could conceivably give rise to incorrect diagnoses and hybridisation to blots of the reaction mixtures was required. Other disadvantages would be that different conditions would have to be determined for each locus of interest which would complicate the simultaneous examination of multiple loci. There would also be the danger of the primers priming at loci other than those desired.

It has not escaped our notice that ARMS may have many other applications in medicine and molecular biology. The technique will be useful for the precise typing of infectious pathogens where characteristic strain-specific base changes can be identified. The analysis of oncogene activation is rendered straightforward as is the detection of deletions in DNA. Many further applications can also be envisaged in the research context.

ACKNOWLEDGEMENTS

We thank Dr D. Holland for the preparation of the oligonucleotides used throughout and all the physicians who provided samples for testing.

Patent applications relating to the methods described here are pending.

REFERENCES

1. Southern, E.M. (1975) J. Mol. Biol. 98 503-517.
2. Kan, Y.W. and Dozy, A.M. (1978) Lancet ii 910-912.
3. Solomon, E. and Bodmer, W.F. (1979) Lancet i 923.
4. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239 487-491.
5. Newton, C.R., Kalsheker, N., Graham, A., Powell, S., Gammack, A., Riley, J. and Markham, A.F. (1988) Nucl. Acid. Res. 16 8233-8243.

Nucleic Acids Research

6. Stoflet, E.S., Koeberl, D.D., Sarkar, G. and Sommer, S.S. (1988) *Science* 239 491-494.
7. Engelke, D.R., Hoener, P.A. and Collins, F.S. (1988) *Proc. Natl. Acad. Sci. USA.* 85 544-548.
8. Bruun Petersen, K., Kølvrå, S., Bolund, L., Bruun Petersen, G., Koch, J. and Gregersen, N. (1988) *Nucl. Acid. Res.* 16 352.
9. Bugawan, T.L., Saiki, R.K., Levenson, C.H., Watson, R.M. and Erlich, H.A. (1988) *Biotechnology* 6 943-947.
10. Chehab, F.F., Doherty, M., Cai, S., Kan, Y.W., Cooper, S. and Rubin, E.M. (1987) *Nature* 329 293-294.
11. Kogan, S.C., Doherty, M. and Gitschier, J. (1987) *New Engl. J. Med* 317 985-990.
12. Levinson, B., Janco, R., Phillips, J. and Gitschier, J. (1987) *Nucl. Acid. Res.* 15 9797-9805.
13. Wong, C., Dowling, C.E., Saiki, R.K., Higuchi, R.G., Erlich, H.A. and Kazazian, H.H. (1987) *Nature* 330 384-386.
14. Connor, B.J., Reyes, A.A., Morin, C. Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983) *Proc. Natl. Acad. Sci. USA.* 80 278-282.
15. Orkin, S.H., Markham, A.F. and Kazazian, H.H. (1983) *J. Clin. Invest.* 71 775-779.
16. Feldman, G., Williamson, R., Beaudet, A.L. and O'Brien, W.E. (1988) *Lancet* ii 102.
17. Williams, C., Williamson, R., Coutelle, C., Loeffler, F., Smith, J. and Ivinson, A. (1988) *Lancet* ii 102-103.
18. Wong, C., Antonarakis, S.E., Goff, S.C., Orkin, S.H., Boehm, C.D. and Kazazian, H.H. (1986) *Proc. Natl. Acad. Sci. USA* 83 6529-6532.
19. Estivill, X., Farrall, M., Scambler, P.J., Bell, G.M., Hawley, K.M.F., Lench, N.J., Bates, G.P., Kruyer, H.C., Frederick, P.A., Stanier, P., Watson, E.K., Williamson, R. and Wainwright, B.J. (1987) *Nature* 326 840-845.
20. Farrall, M., Estivill, X. and Williamson, R. (1987) *Lancet* ii 155-157.
21. Nukiwa, T., Brantly, M., Garver, R., Paul, L., Courtney, M., LeCocq, J.P. and Crystal, R.G. (1986) *J. Clin. Invest.* 77 528-537.
22. Ludwig, E.H., Blackhart, B.D., Pierotti, V.R., Caiati, L., Fortier, C., Knott, T., Scott, J., Mahley, R.W., Levy-Wilson, B. and McCarthy, B.J. (1987) *DNA* 6 363-372.
23. Chien, A., Edgar, D.B. and Trela, J.M. (1976) *J. Bacteriol.* 127 1550-1557.
24. Long, G.L., Chandra, T., Woo, S.L.C., Davie, E.W. and Kurachi, K. (1984) *Biochemistry* 23 4828-4837.
25. Tindall, K.R. and Kunkel, T.A. (1988) *Biochemistry* 27 6008-6013.
26. Cooper, D.N. and Youssoufian, H. (1988) *Hum. Genet.* 78 151-155.
27. Dermer, S.J. and Johnson, E.M. (1988) *Lab. Invest.* 59 403-408.

EXHIBIT 6

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE**

ARCHERDX, INC. and THE GENERAL
HOSPITAL CORPORATION d/b/a
MASSACHUSETTS GENERAL
HOSPITAL,

Plaintiffs,

v.

QIAGEN SCIENCES, LLC, QIAGEN LLC
f/k/a QIAGEN, INC., QIAGEN BEVERLY,
INC., QIAGEN GAITHERSBURG, INC.,
QIAGEN GMBH, QIAGEN N.V., and
JONATHAN ARNOLD,

Defendants.

Civil Action No. 18-cv-01019-MN-CJB

**HIGHLY CONFIDENTIAL –
ATTORNEYS’ EYES ONLY**

**REPLY EXPERT REPORT OF DR. MICHAEL METZKER REGARDING
THE INVALIDITY OF U.S. PATENT NOS. 10,017,810 AND 10,450,597**

TABLE OF CONTENTS

I.	QUALIFICATIONS/COMPENSATION/MATERIALS REVIEWED	1
A.	Education/Training/Expertise	1
B.	Compensation	1
C.	Materials Reviewed	2
II.	TECHNICAL BACKGROUND.....	2
III.	APPLICABLE STANDARDS	2
A.	Obviousness Standards	3
B.	Written Description and Indefiniteness Standards.....	4
C.	The Meaning of Claim Terms and Scope of the Asserted Claims to One of Ordinary Skill in the Art	5
IV.	THE PATENTS IN SUIT AND ASSERTED CLAIMS	7
A.	Patent Priority of the Asserted Claims of the Patents-in-Suit.....	7
V.	INVALIDITY OF U.S. PATENT NO. 10,017,810 (Claims 16-20).....	7
A.	The Asserted Claims Are Invalid for Lack of Written Description.....	8
1.	Non-Nested Second Target-Specific Primer.....	8
2.	Second Target-Specific Primer Annealing to Sequence Other Than DNA to which Adaptors Are Ligated.....	14
3.	Sequence Identical to a Second Sequencing Primer	15
4.	Sequencing Primers in the Universal Oligonucleotide Tail Adaptor.....	16
5.	Second Adaptor Primer.....	18
B.	The Asserted Claims Are Invalid As Indefinite.....	19
C.	The Asserted Claims Are Anticipated and Obvious	21
1.	The Prior Art Discloses the Claimed Adaptor or Renders the Claimed Invention Obvious.....	22
2.	Amsterdam Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	26
3.	Siebert Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	34
4.	Markham Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	39
5.	Chenchik Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	49

VI.	INVALIDITY OF U.S. PATENT NO. 10,450,597 (Claims 1, 3-5, 10, 11, 13-27, and 29)	54
A.	The Asserted Claims Are Invalid for Lack of Written Description.....	55
B.	The Asserted Claims Are Invalid As Indefinite.....	58
C.	The Asserted Claims Are Anticipated and Obvious.....	58
1.	The Degenerate Primer References Do Not Anticipate or Render Obvious the Claimed Invention of the '597 Patent	60
2.	Fraiture Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	65
3.	McCarty Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	69
4.	Morley Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	71
5.	Martin-Harris Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	74
6.	The APAGene Manual Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	78
7.	'810 References	81
VII.	XII. Secondary Considerations	81
A.	Commercial Success	82
B.	Industry Praise	83
C.	Long-Felt Need	84
D.	Skepticism.....	85
E.	Simultaneous Invention	86
I.	QUALIFICATIONS/COMPENSATION/MATERIALS REVIEWED.....	1
A.	Education/Training/Expertise	1
B.	Compensation	1
C.	Materials Reviewed	2
II.	TECHNICAL BACKGROUND.....	2
III.	APPLICABLE STANDARDS	2
A.	Obviousness Standards	3
B.	Written Description and Indefiniteness Standards.....	4
C.	The Meaning of Claim Terms and Scope of the Asserted Claims to One of	

	Ordinary Skill in the Art	5
IV.	THE PATENTS IN SUIT AND ASSERTED CLAIMS	7
A.	Patent Priority of the Asserted Claims of the Patents-in-Suit.....	7
V.	INVALIDITY OF U.S. PATENT NO. 10,017,810 (Claims 16-20).....	7
A.	The Asserted Claims Are Invalid for Lack of Written Description.....	8
1.	Non-Nested Second Target-Specific Primer.....	8
2.	Second Target-Specific Primer Annealing to Sequence Other Than DNA to which Adaptors Are Ligated.....	14
3.	Sequence Identical to a Second Sequencing Primer	15
4.	Sequencing Primers in the Universal Oligonucleotide Tail Adaptor.....	16
5.	Second Adaptor Primer.....	18
B.	The Asserted Claims Are Invalid As Indefinite.....	19
C.	The Asserted Claims Are Anticipated and Obvious	21
1.	The Prior Art Discloses the Claimed Adaptor or Renders the Claimed Invention Obvious.....	22
2.	Amsterdam Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	26
a)	Claim 16[a]: Step (i)	26
b)	Claim 16[c]: Step (iii)	33
c)	Dependent Claims.....	34
3.	Siebert Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	34
a)	Claim 16[a]: Step (i)	35
b)	Dependent Claims.....	39
4.	Markham Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	39
a)	Claim 16[a]: Step (i)	39
b)	Dependent Claims.....	48
5.	Chenchik Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	49
a)	Claim 16[a]: Step (i)	49
b)	Claim 16[e]: Barcode.....	53
c)	Dependent Claims.....	54
VI.	INVALIDITY OF U.S. PATENT NO. 10,450,597 (Claims 1, 3-5, 10, 11, 13-27, and 29)	54

A.	The Asserted Claims Are Invalid for Lack of Written Description.....	55
B.	The Asserted Claims Are Invalid As Indefinite.....	58
C.	The Asserted Claims Are Anticipated and Obvious	58
1.	The Degenerate Primer References Do Not Anticipate or Render Obvious the Claimed Invention of the '597 Patent	60
2.	Fraiture Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	65
a)	Claim 1[b]: A Plurality of Different Primers that Share a Common Sequence that Is 5' to Different Hybridization Sequences	65
b)	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends.....	66
c)	Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence	67
d)	Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence	68
e)	Dependent Claims	69
3.	McCarty Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	69
a)	Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence	69
b)	Dependent Claims	70
4.	Morley Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	71
a)	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends.....	71
b)	[Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence].....	72
c)	[Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence].....	73
d)	Dependent Claims	74
5.	Martin-Harris Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	74
a)	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends.....	74
b)	[Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence].....	76

c)	[Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence].....	77
d)	Dependent Claims	78
6.	The APAGene Manual Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art	78
a)	Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends.....	78
b)	[Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence].....	79
c)	[Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence].....	80
d)	Dependent Claims	81
7.	'810 References	81
VII.	XII. Secondary Considerations	81
A.	Commercial Success	82
B.	Industry Praise	83
C.	Long-Felt Need	84
D.	Skepticism.....	85
E.	Simultaneous Invention	86

B. The Asserted Claims Are Invalid As Indefinite

129. Dr. Lennon contends that even when step (a) and step (b) of claim 1 of the '597 Patent are performed out of order, that is, step (b) is performed first and then step (a), one of ordinary skill in the art would have understood that the phrase *the first strand of the double-stranded target nucleic acid* in step (b) to be a *first strand of the double-stranded target nucleic acid* strand in step (a).²²⁶ I disagree as there is no antecedent basis of the phrase *the first strand* if step (b) is performed first, and Dr. Lennon provides no explanation as to how one of ordinary skill in the art would have interpreted the claim other than “[t]here is only one ‘first strand’ described in claim 1.”²²⁷

130. Dr. Lennon also does not shed any additional light on the ambiguous phrase “the same sequence of the double stranded target nucleic acid” in step (b) of claim 1 of the '597 Patent other than by the statement that it does not “matter which sequence it may be.”²²⁸

131. Dr. Lennon contends that my “analysis of the prior art also confirms that the terms ‘the first strand’ and ‘the same sequence, do not render the '597 claims indefinite.”²²⁹ To the extent I understand Dr. Lennon’s argument, I have been informed and understand that a patent may be invalid for indefiniteness and separately invalid as anticipated or obvious.²³⁰

C. The Asserted Claims Are Anticipated and Obvious

132. After reviewing the Lennon Rebuttal Report, my opinion has not changed that the following prior art references anticipate or render obvious the Asserted Claims of the '597 Patent.

²²⁶ See *id.* at ¶ 315.

²²⁷ See *id.*

²²⁸ See *id.* at ¶ 316.

²²⁹ See *id.* at ¶ 317.

²³⁰ See *supra* ¶ 14.

- Fraiture²³¹ anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art.²³²
- McCarty²³³ anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art.²³⁴
- Morley²³⁵ anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art.²³⁶
- Martin-Harris²³⁷ anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art.²³⁸
- The APAGene Manual²³⁹ anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art.²⁴⁰
- The '810 and '828 Patents²⁴¹ anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art.²⁴²

133. Below, I address certain points raised by Dr. Lennon. It should be understood, however, that others points not specifically addressed here should not be inferred as unspoken

²³¹ See Metzker Opening Report Ex. QQ, QIA-00689151-00689159, Fraiture *et al.*, *An innovative and integrated approach based on DNA walking to identify unauthorized GMOs*, FOOD CHEMISTRY 147:60-69 (2014) ("Fraiture").

²³² See Metzker Opening Report at ¶¶ 304-376.

²³³ See Metzker Opening Report Ex. SS, QIA-00689182-00689192, McCarty *et al.*, *Mu-seq: Sequence-based mapping and identification of transposon induced mutations*, PLOS ONE 8:e77172 pp. 1-11 (2013) ("McCarty").

²³⁴ See Metzker Opening Report at ¶¶ 377-455.

²³⁵ See Metzker Opening Report Ex. RR, QIA-00689264-00689363, Morley, *Method of DNA amplification*, UNITED STATES PATENT APPLICATION NO. US 2010/0248310 (2010) ("Morley").

²³⁶ See Metzker Opening Report at ¶¶ 456-544.

²³⁷ See Metzker Opening Report Ex. PP, QIA-00689179-00689181, Martin-Harris *et al.*, *Gene walking using sequential hybrid primer polymerase chain reaction*, ANALYTICAL CHEMISTRY 399:308-310 (2010) ("Martin-Harris").

²³⁸ See Metzker Opening Report at ¶¶ 545-635.

²³⁹ See Metzker Opening Report Ex. TT, QIA-00689102-00689115, Bio S&T Inc., *APAGene™ GOLD-RT genome walking kits* (2010) ("APAGene Manual").

²⁴⁰ See Metzker Opening Report at ¶¶ 636-731.

²⁴¹ See Metzker Opening Report Ex. CCC, QIA-00689213-00689263, '810 Patent; *see also* Metzker Opening Report Ex. DDD, QIA-00689416-00689468, Iafrate *et al.*, *Methods for determining a nucleotide sequence contiguous to a known target nucleotide sequence*, UNITED STATES PATENT NO. 9,487,828 ("'828 Patent").

²⁴² See Metzker Opening Report at ¶¶ 732-788.

agreements with the opinions of Dr. Lennon. To the contrary, I believe that many of the specific points raised in the Lennon Rebuttal Report that are not addressed below were adequately addressed in my Opening Report.

1. The Degenerate Primer References Do Not Anticipate or Render Obvious the Claimed Invention of the '597 Patent

134. Dr. Lennon relies on a encyclopedia book chapter that I published in 2006.²⁴³ Dr. Lennon mischaracterizes the section titled, *Degenerate polymerase chain reaction*, to suggest that I teach that “[l]owering the PCR annealing temperature increases the likelihood that a degenerate primer will anneal to a region of interest, even with less than 100% complementarity (i.e., even with some degree of mismatched bases).”²⁴⁴ This chapter makes no such statements. What I do teach is how to minimize the complexity of a pool of degenerate primers by using codon bias and the use conserved sequences across species.

Degenerate PCR is a powerful strategy for obtaining novel full-length cDNA sequences from limited amino acid sequence information. The PCR primer sequences are derived from the reverse translation of 6-9 amino acid codons, which will result in varying levels of degeneracy except for methionine and tryptophan residues. Careful attention should be exercised in the design of degenerate primers because increasing the primer complexity (i.e., using codons that show more than twofold degeneracy) will typically result in an increase in nonspecific PCR products. ***One approach in reducing the complexity of the degenerate primer*** is the use of codon bias for the particular organism from which the gene will be cloned. Alternatively, ***the alignment of orthologous gene sequences from other species can greatly improve the specificity*** of cloning the gene of interest by revealing evolutionarily conserved domains.²⁴⁵

²⁴³ See Metzker Opening Report Ex. H, Metzker and Caskey, *Polymerase chain reaction*, ENCYCLOPEDIA OF MEDICAL DEVICES AND INSTRUMENTATION, SECOND EDITION, Volume 5, Ed. J.G. Webster (2006) (“Metzker 2006 Chapter”). I note that Dr. Lennon refers this chapter as “his 2001 article,” which is incorrect.

²⁴⁴ See Lennon Rebuttal Report at ¶ 320.

²⁴⁵ See Metzker 2006 Chapter at 385 (emphases added, citations omitted).

One of ordinary skill in the art would have understood that reducing the complexity of the pool of degenerate primers means that “not all possible permutations” exist in the sequences of the degenerate primers, which is unlike Dr. Lennon’s suggestion.²⁴⁶ The outcome of carefully designing the pool of degenerate primers sequences is so that **“[t]he specificity of PCR should then selectively amplify the correct primer sequences** to generate a gene or gene family-specific probe from which the full length cDNA can be obtained.”²⁴⁷ In other words, a specific sequence within the pool of degenerate primers will correctly bind or anneal to the target of interest to amplify to desired product. Thus, Dr. Lennon’s contention of lowering the temperature to promote mismatching is misleading.²⁴⁸

135. Dr. Lennon’s assertion that annealing temperatures in the range of “55°C to 68°C”²⁴⁹ is also misleading as I teach that such annealing conditions are “. . . dependent on the concentration, base composition and the length of the oligonucleotide.”²⁵⁰ Concentration is of particular importance as the amount of a given sequence in the degenerate pool is typically very low. To compensate for such low concentrations, one of ordinary skill in the art would have understood that lowering the annealing temperature would increase the probability of the binding or annealing of a given sequence to its complementary sequence. The trade-off here, unfortunately, is that lower temperatures inadvertently increases the amount of non-specific products during amplification.

²⁴⁶ See Lennon Rebuttal Report at ¶ 320.

²⁴⁷ See Metzker 2006 Chapter at 385 (emphasis added).

²⁴⁸ See Lennon Rebuttal Report at ¶ 320.

²⁴⁹ See *id.* at ¶ 321.

²⁵⁰ See Metzker 2006 Chapter at 381.

136. I also teach that “[t]he use of the cosolvents glycerol and dimethyl sulfoxide (DMSO) have been shown to protect against DNA damage by efficiently lowering the denaturation temperature by several degrees centigrade.”²⁵¹ These solvents also lower the annealing temperature. Thus, making comparisons of general reaction conditions based on specific examples in the literature without accounting for any relevant differences is a non-productive exercise and can lead to incorrect conclusions.

137. Dr. Lennon does nothing more than speculate that the strategy of using degenerate primers is to promote mismatching between sequences to increase the likelihood of amplifying a region of interest.

In a degenerate primer strategy, such mismatches are not necessarily an undesirable outcome for a researcher because mismatches increase the likelihood that any given region of interest will be amplified. Lowering the annealing temperature can increase the likelihood that at least some of the degenerate primers will amplify the region of interest even ***if none of the degenerate primers have 100% complementarity with the region of interest.***²⁵²

I disagree, and as described above, careful attention should be exercised in the design of degenerate primers so that the correct region of interest is amplified.

138. Dr. Lennon also conflates the inadvertent amplification of non-specific products with his intentional approach to promote mismatching in PCR. He then relies on this incorrect approach to suggest that nested PCR can be used to “‘clean up’ the degenerate primer products.”²⁵³ Dr. Lennon cites to Morley to contend that the data in Figure 5 “ . . . illustrates how a first round of PCR with degenerate primers run at low temperatures (first 35°C, then 55°C) produces a greater

²⁵¹ See *id.* at 383.

²⁵² See Lennon Rebuttal Report at ¶ 322 (emphases added). I note that Dr. Lennon provides no scientific literature to support this position.

²⁵³ See *id.* at ¶ 323.

abundance of non-specific products (illustrated as being detectable with fewer cycles [Ct]) than specific product.”²⁵⁴ I disagree with this analysis of Morley for several reasons. First, the lower annealing temperature of 35°C in the first five cycles was most likely used because specific primer sequences in the degenerate pool are found in low concentrations, as described above. Lower annealing temperatures are used to increase the probability of binding or annealing of a given primer sequence generally, which is especially important when there is a low concentration of specific primer sequences. Secondly, Morley’s specification does not associate the conditions cited above in Example 3 with the data shown in Figure 5. For example, the PCR products created in Example 3 were “ . . . examined by gel electrophoresis and discrete bands were isolated and sequenced”²⁵⁵ The data shown in Figure 5, however, appear to be the results of real-time PCR analysis. Therefore, Figure 5 includes results that were not generated by the conditions from Example 3.

139. Dr. Lennon contends that “ . . . [t]he low-temperature annealing conditions of [my] five prior art references and the use of degenerate primers (with some highly degenerate) would be expected to produce significant mismatches.”²⁵⁶ I disagree for several reasons. As described above, lower annealing temperatures are required to increase the probability of correct binding or annealing of a given primer at low concentration. Any incorrect binding or annealing of a primer would be an unintentional event, and most likely would result in off-target binding events (*i.e.*, not within the target nucleic acid). Unlike Dr. Lennon’s suggestion, the purpose of operating at lower annealing temperatures is not to “produce significant mismatches” during primer binding or annealing to off-target regions.²⁵⁷

²⁵⁴ See *id.*

²⁵⁵ See Morley at [0252].

²⁵⁶ See Lennon Rebuttal Report at ¶ 324.

²⁵⁷ See *id.*

140. Dr. Lennon attempts to construe the phrase “. . . each primer of the plurality of different primers **does not anneal** to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers,”²⁵⁸ which has not been construed by the Court.²⁵⁹ I note that Dr. Lennon relies on the terms “anneal” and “complementary” to essentially arrive at an overtly narrow construction that “broadly excludes different primers that anneal to the same sequence, including annealing where there is less than 100% complementarity between a primer and a template.”²⁶⁰ I disagree and point out that one term not considered by Dr. Lennon is *target nucleic acid*, which the Court has construed to mean “a nucleic acid molecule of interest (e.g., a nucleic acid to be analyzed).”²⁶¹ Dr. Lennon’s construction would include the entire genome and exclude all off-target binding or annealing events as not *meeting the negative limitation of step (b)*. What Dr. Lennon and Dr. McSwiggen overlook are the mispriming events that occur only within the *target nucleic acid*. In my opinion, the careful design of degenerate primers as taught in the prior art would meet the *negative limitation of step (b)*. For the reasons described above, I disagree with Dr. Lennon’s contention that I fail “. . . to establish by clear and convincing evidence that any of the degenerate primers in Fraiture, McCarty, Morley, Martin-Harris and the APAGene Manual **do not anneal** to the same sequence.”²⁶² In the following sections, I provide my analysis of specific prior art references in response to those raised by Dr. Lennon.

²⁵⁸ See *id.* at ¶ 325. Dr. Lennon refers to this phrase as the “negative limitation in step (b).” of claim 1 of the ’597 Patent.

²⁵⁹ See D.I. 254, *Memorandum Order* at p. 1.

²⁶⁰ See Lennon Rebuttal Report at ¶ 327.

²⁶¹ See D.I. 254, *Memorandum Order* at p. 1.

²⁶² See Lennon Rebuttal Report at ¶ 329.

2. Fraiture Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art

141. After reviewing the Lennon Rebuttal Report, my opinion has not changed that Fraiture anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art references. Below, I address specific points raised by Dr. Lennon in his Rebuttal Report.

a) Claim 1[b]: A Plurality of Different Primers that Share a Common Sequence that Is 5' to Different Hybridization Sequences

142. Dr. Lennon attempts to argue that any differences between the APAGene Manual and APAGene kit described in Fraiture are material. I disagree. For example, Dr. Lennon contends that the letters “RT” are not included in the Fraiture reference.

There are differences between the APAGene Manual and the APAGene kit described in Fraiture. For example, *the APAGene Manual bears the title “APAGene™ Gold-RT Genome Walking Kits”* with “Cat. # BT901-RT” whereas the kit described in *Fraiture does not include “RT” in the title* and a catalog number is not provided.²⁶³

While true that APAGene Manual is titled *APAGene™ Gold-RT Genome Walking Kits*, the description below the title refers to the kits as the *APAGene™ Gold Genome Walking Kits*,²⁶⁴ which does not have the “RT” letters and is exactly the same title cited in Fraiture.²⁶⁵ Therefore, there is no question that the reagents used in Fraiture are disclosed in the APAGene Manual.²⁶⁶

143. Dr. Lennon also draws attention to the “. . . differences between the primer sequences in the APAGene Manual and the sequences and schematic disclosed in Fraiture.”²⁶⁷ While I agree that Figure 1b in Fraiture could be drawn more precisely, the UAP-N2 is shown

²⁶³ See *id.* at ¶ 333 (emphases added).

²⁶⁴ See APAGene Manual at p. 1

²⁶⁵ See Fraiture at 65.

²⁶⁶ This conclusion addresses Dr. Lennon’s contention in his Rebuttal Report at ¶ 339.

²⁶⁷ See Lennon Rebuttal Report at ¶ 334.

slightly to the right, which is consistent with the sequences shown in ¶ 335 of the Lennon Rebuttal Report. Therefore, any difference between the sequences and schematic in Fraiture are immaterial.

b) Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends

144. Dr. Lennon contends that “[b]ecause the 3’ ends of each DRT primer are the same, Dr. Metzker fails to demonstrate that the plurality of different primers have hybridization sequences with different 3’ ends.”²⁶⁸ I disagree as this is not the manner the prior art describes different 3’-ends. For example, Newton 1989²⁶⁹ described an improved PCR method that utilized mismatched nucleotides on or near the 3’-end of primers resulting in different 3’-ends. Newton 1989 claimed that “[t]he basis of the invention is that unexpectedly, **oligonucleotides with a mismatched 3’-residue** will not function as primers in the PCR under appropriate conditions.”²⁷⁰ Newton 1989 observed that the mismatches did not have to occur directly at the 3’-end but can also occur internal to the 3’-end and still function as primers having different 3’-ends.

In some instances a single 3’-mismatched base does allow amplification to proceed. ***We have shown that introducing additional deliberate mismatches near the 3’ end of appropriate primers ameliorates this problem.***²⁷¹

Newton 1989 demonstrates internal mismatches up to seven bases from the 3’-end with improved specificity with primers having different 3’-ends.

In an attempt to increase the specificity of the ARMS primers ***we chose to deliberately introduce an additional mismatch near their 3’-ends.*** When primers 7f and 7g which have ***a deliberate C/T mismatch seven bases from***

²⁶⁸ See Lennon Rebuttal Report at ¶ 340.

²⁶⁹ See Newton *et al.* *Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS).* NUCLEIC ACIDS RESEARCH 17:2503-2516 (1989) (“Newton 1989”).

²⁷⁰ See *id.* at Abstract (emphasis added).

²⁷¹ See *id.* at 2504 (emphasis added).

the 3'-end were introduced to replace primers 7 and 7a (figure 2) specificity was improved.²⁷²

One of ordinary skill in the art would have understood that the primers in Newton 1989 having different 3'-end were analogous to the DRT primers using in Fraiture and the APAGene Manual. Therefore, I disagree with Dr. Lennon's conclusion that "Fraiture and the APAGene Manual fail to disclose a plurality of different primers that have different hybridization sequences having different 3' ends,"²⁷³ for the reasons described above.

c) Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence

145. As described above, there is no question that the reagents used in Fraiture are disclosed in the APAGene Manual.²⁷⁴ I have reviewed the McSwiggen Report.²⁷⁵ Dr. Lennon also "reviewed the expert report of Dr. McSwiggen and agree with his methods, analysis and conclusion."

Under the permissive PCR conditions specified in APAGene Manual (i.e., annealing temperature equal to 42°C), and given the high level of degeneracy of the DRT Primers, ***a POSA would have expected at least two different primers in each DRT Primer mix to anneal to the same sequence of template DNA.***²⁷⁶

I disagree with the McSwiggen Report as it did not consider the Court's construction of the term *target nucleic acid* in his analysis of the DRT primers.²⁷⁷ As described above, the Court's construction for the term *target nucleic acid* is a "nucleic acid molecule of interest (e.g., a nucleic

²⁷² See *id.* at 2511 (emphases added).

²⁷³ See Lennon Rebuttal Report at ¶ 343.

²⁷⁴ This conclusion addresses Dr. Lennon's contention in his Rebuttal Report at ¶ 345.

²⁷⁵ See Expert Report of James A. McSwiggen, Ph.D, dated November 20, 2020 ("McSwiggen Report")

²⁷⁶ See Lennon Rebuttal Report at ¶ 347 (emphasis added).

²⁷⁷ See McSwiggen Report at ¶¶ 9-10, purporting to rely on the Court's construed terms "plurality of different primers," "the same sequence," "double-stranded target nucleic acid," and the '597 Patent definitions "anneal" and "complementary."

acid to be analyzed).” In Fraiture, the author’s strategy was to find local regions in the rice genome that contain pCAMBIA vector insertions.²⁷⁸ Based on their approach, Fraiture reported target nucleic acids in the range of 300 to 1,000 base pairs.²⁷⁹ Dr. McSwiggen, however, considered the entire rice genome,²⁸⁰ which is approximately 420 megabase pairs,²⁸¹ which is roughly 420,000 times larger than the target nucleic acids of Fraiture, and most, if not all, of the hits would be expected to be off-target events. Thus, the analysis provided by Dr. McSwiggen and relied on by Dr. Lennon is flawed and should not be afforded any weight.

d) Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence

146. Dr. Lennon argues that “no portion of the t35S pCAMBIA b-R primer anneals to any portion of the complement of the target-specific hybridization sequence of the t35S pCAMBIA a-R primer” based on an incorrect application of the Court’s claim construction for “target-specific hybridization sequence” to this limitation. The limitation in question for Claim 1[e] is not “target-specific hybridization sequence” but rather the “complement of the target-specific hybridization sequence.” The Court did not construe “complement of the target-specific hybridization sequence” and Dr. Lennon has provided no basis for his argument that a primer that specifically anneals to the “complement of the target-specific hybridization sequence” must specifically bind to the portion of the complement that is complementary to the sequence of the “target-specific primer.” This narrow interpretation is inconsistent with his broad interpretations of other claim terms as discussed throughout this report and in my Opening Report. As discussed in paragraph 319 of my

²⁷⁸ See Fraiture at Abstract.

²⁷⁹ See *id.* at 67.

²⁸⁰ See McSwiggen Report at ¶ 28.

²⁸¹ See Ex. MMM, Yu *et al.*, *Draft sequence of the rice genome (Oryza sativa L. ssp. Indica)*, SCIENCE 296:79-92 (2002) (“Yu”) at Abstract.

Opening Report, the t35S pCAMBIA b-R primer specifically anneals to the “complement of the target-specific hybridization sequence” and Dr. Lennon has failed to point to any construction by the Court that is inconsistent with the disclosure. For at least the foregoing reasons, his analysis is flawed and should not be afforded any weight.

e) Dependent Claims

147. Dr. Lennon argues that I failed “to establish that dependent claims 3-5, 10, 11, 13-27, and 29 are anticipated or obvious”²⁸² based on my his analysis of claim 1 of the ’597 Patent. As described above, Dr. Lennon’s assertions are incorrect, and Fraiture anticipates claim 1 of the ’597 Patent or renders it obvious in combination with the prior art. I also incorporate my reasons why Markham anticipates claims 3-5, 10, 11, 13-27, and 29 of the ’597 Patent or renders them obvious in combination with the prior art.²⁸³

3. McCarty Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art

148. After reviewing the Lennon Rebuttal Report, my opinion has not changed that McCarty anticipates the Asserted Claims of the ’597 Patent or renders them obvious in combination with other prior art references. Below, I address specific points raised by Dr. Lennon in his Rebuttal Report.

a) Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence

149. As described above, I have reviewed the McSwiggen Report. Dr. Lennon also “reviewed the expert report of Dr. McSwiggen and agree with his methods, analysis and conclusion.”

²⁸² See Lennon Rebuttal Report ¶ 357.

²⁸³ See Metzker Opening Report ¶¶ 320-376

As indicated in Table 5 of his report, reproduced below, the predicted Tms for each of the 12 Mu-TIR primers ranges from 67.1°C to 75.4°C degrees, which are significantly higher than either of the two annealing temperatures (54°C and 60°C) that McCarty prescribes for PCR I.²⁸⁴

I disagree with the McSwiggen Report as it did not consider the Courts' construction of the term *target nucleic acid* in his analysis of the DRT primers.²⁸⁵ As described above, the Court's construction for the term *target nucleic acid* is a "nucleic acid molecule of interest (e.g., a nucleic acid to be analyzed)." In fact, Dr. Lennon noted that "McCarty does not provide sufficient information to allow Dr. McSwiggen to identify the specific strains of maize DNA, he used the reverse complement of each Mu-TIR primer as the target."²⁸⁶ Moreover, the McSwiggen Report only estimated Tms for the primers.²⁸⁷ Nonetheless, Dr. Lennon concludes from this incomplete data set that "a POSA would have understood that at least some quantity of two or more primers from the degenerate Mu-TIR primers disclosed in McCarty would have annealed to the same target sequence."²⁸⁸ Thus, the analysis provided by Dr. McSwiggen and relied on by Dr. Lennon is flawed and should not be afforded any weight.

b) Dependent Claims

150. Dr. Lennon argues that I failed "to establish that dependent claims 3-5, 10, 11, 13-27, and 29 are anticipated or obvious"²⁸⁹ based on my his analysis of claim 1 of the '597 Patent. As described above, Dr. Lennon's assertions are incorrect, and Fraiture anticipates claim 1 of the

²⁸⁴ See Lennon Rebuttal Report at ¶ 347 (emphasis added).

²⁸⁵ See McSwiggen Report at ¶¶ 9-10, purporting to rely on the Court's construed terms "plurality of different primers," "the same sequence," "double-stranded target nucleic acid," and the '597 Patent definitions "anneal" and "complementary."

²⁸⁶ See Lennon Rebuttal Report ¶ 362.

²⁸⁷ See McSwiggen Report at ¶ 56.

²⁸⁸ See Lennon Rebuttal Report ¶ 363.

²⁸⁹ See Lennon Rebuttal Report ¶ 364.

'597 Patent or renders it obvious in combination with the prior art. I also incorporate my reasons why Markham anticipates claims 3-5, 10, 11, 13-27, and 29 of the '597 Patent or renders them obvious in combination with the prior art.²⁹⁰

4. Morley Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art

151. After reviewing the Lennon Rebuttal Report, my opinion has not changed that Morley anticipates the Asserted Claims of the '597 Patent or renders them obvious in combination with other prior art references. Below, I address specific points raised by Dr. Lennon in his Rebuttal Report.

a) Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3' Ends

152. Dr. Lennon contends that I “highlight[] in magenta is the 5' end of the hybridization sequence rather than the 3' end, as the claim requires.”²⁹¹ I disagree as this is not the manner the prior art describes different 3'-ends. For example, Newton 1989 described an improved PCR method that utilized mismatched nucleotides on or near the 3'-end of primers resulting in different 3'-ends. Newton 1989 claimed that “[t]he basis of the invention is that unexpectedly, *oligonucleotides with a mismatched 3'-residue* will not function as primers in the PCR under appropriate conditions.”²⁹² Newton 1989 observed that the mismatches did not have to occur directly at the 3'-end but can also occur internal to the 3'-end and still function as primers having different 3'-ends.

In some instances a single 3'-mismatched base does allow amplification to proceed. *We have shown that introducing additional deliberate*

²⁹⁰ See Metzker Opening Report ¶¶ 398-455.

²⁹¹ See Lennon Rebuttal Report at ¶ 369.

²⁹² See Newton 1989at Abstract (emphasis added).

*mismatches near the 3' end of appropriate primers ameliorates this problem.*²⁹³

Newton 1989 demonstrates internal mismatches up to seven bases from the 3'-end with improved specificity with primers having different 3'-ends.

In an attempt to increase the specificity of the ARMS primers *we chose to deliberately introduce an additional mismatch near their 3'-ends*. When primers 7f and 7g which have *a deliberate C/T mismatch seven bases from the 3'-end* were introduced to replace primers 7 and 7a (figure 2) specificity was improved.²⁹⁴

One of ordinary skill in the art would have understood that the primers in Newton 1989 having different 3'-end were analogous to the degenerate primers using in Morley. Therefore, I disagree with Dr. Lennon's conclusion that "Morley fails to disclose a plurality of different primers that have different hybridization sequences having different 3' ends,"²⁹⁵ for the reasons described above.

b) [Claim 1(c): Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence]

153. As described above, I have reviewed the McSwiggen Report. Dr. Lennon also "reviewed the expert report of Dr. McSwiggen and agree with his methods, analysis and conclusion."²⁹⁶

As Dr. McSwiggen explains, his analysis for Morley is identical to that of Martin-Harris because Morley's objective, target genes, primer sets, and reaction conditions are the same, except that Morley discloses only one degenerate primer (labeled "GW-A" by Dr. McSwiggen) of the four gene walking degenerate primer types disclosed in Martin-Harris.²⁹⁷

154. As Dr. McSwiggen explains his analysis of Morley is identical to that of Martin-

²⁹³ See *id.* at 2504 (emphasis added).

²⁹⁴ See *id.* at 2511 (emphases added).

²⁹⁵ See Lennon Rebuttal Report at ¶ 369.

²⁹⁶ See *id.* at ¶ 371.

²⁹⁷ See *id.*

Harris, I incorporate my comments regarding the Martin-Harris analysis, *see infra* at ¶ 158.

c) [Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence]

155. Dr. Lennon argues that “Dr. Metzker does not identify any nested target-specific primer sequences for the three genes analyzed in Example 3, and therefore fails to demonstrate that the F2 of Example 3 anneals to the complement of F1, as the ’597 claim requires” based on an incorrect application of the Court’s claim construction for “target-specific hybridization sequence” to this limitation.²⁹⁸ The limitation in question for Claim 1[e] is not “target-specific hybridization sequence” but rather the “complement of the target-specific hybridization sequence.” The Court did not construe “complement of the target-specific hybridization sequence” and Dr. Lennon has provided no basis for his argument that a primer that specifically anneals to the “complement of the target-specific hybridization sequence” must specifically bind to the portion of the complement that is complementary to the sequence of the “target-specific primer.” This narrow interpretation is inconsistent with his broad interpretations of other claim terms as discussed throughout this report and in my Opening Report. As discussed in paragraphs 468-470 of my Opening Report, Morley discloses subjecting the extension product from (b) to a first PCR amplification reaction using a tail primer (hybrid reverse primer, shown as “tag1”) comprising a 3’ sequence that specifically anneals to the complement of the common sequence and a 5’ tail sequence and a nested target-specific primer (F2) that specifically anneals to the complement of the target-specific hybridization sequence. Dr. Lennon has failed to point to any construction by the Court that is inconsistent with this disclosure. For at least the foregoing reasons, his analysis is flawed and should not be afforded any weight.

²⁹⁸ See Lennon Rebuttal Report at ¶¶ 373-375.

d) Dependent Claims

156. Dr. Lennon argues that I failed “to establish that dependent claims 3-5, 10, 11, 13-27, and 29 are anticipated or obvious”²⁹⁹ based on my his analysis of claim 1 of the ’597 Patent. As described above, Dr. Lennon’s assertions are incorrect, and Fraiture anticipates claim 1 of the ’597 Patent or renders it obvious in combination with the prior art. I also incorporate my reasons why Markham anticipates claims 3-5, 10, 11, 13-27, and 29 of the ’597 Patent or renders them obvious in combination with the prior art.³⁰⁰

5. Martin-Harris Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art

157. After reviewing the Lennon Rebuttal Report, my opinion has not changed that Martin-Harris anticipates the Asserted Claims of the ’597 Patent or renders them obvious in combination with other prior art references. Below, I address specific points raised by Dr. Lennon in his Rebuttal Report.

a) Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3’ Ends

158. Dr. Lennon contends that “[e]ach of the four different degenerate primers have identical 3’ ends, as illustrated above, and expressly stated in Martin-Harris.”³⁰¹ I disagree as this is not the manner the prior art describes different 3’-ends. For example, Newton 1989 described an improved PCR method that utilized mismatched nucleotides on or near the 3’-end of primers resulting in different 3’-ends. Newton 1989 claimed that “[t]he basis of the invention is that unexpectedly, *oligonucleotides with a mismatched 3’-residue* will not function as primers in the

²⁹⁹ See Lennon Rebuttal Report ¶ 376.

³⁰⁰ See Metzker Opening Report ¶¶ 471-544.

³⁰¹ See Lennon Rebuttal Report at ¶ 382.

PCR under appropriate conditions.”³⁰² Newton 1989 observed that the mismatches did not have to occur directly at the 3’-end but can also occur internal to the 3’-end and still function as primers having different 3’-ends.

In some instances a single 3’-mismatched base does allow amplification to proceed. ***We have shown that introducing additional deliberate mismatches near the 3’ end of appropriate primers ameliorates this problem.***³⁰³

Newton 1989 demonstrates internal mismatches up to seven bases from the 3’-end with improved specificity with primers having different 3’-ends.

In an attempt to increase the specificity of the ARMS primers ***we chose to deliberately introduce an additional mismatch near their 3’-ends.*** When primers 7f and 7g which have ***a deliberate C/T mismatch seven bases from the 3’-end*** were introduced to replace primers 7 and 7a (figure 2) specificity was improved.³⁰⁴

One of ordinary skill in the art would have understood that the primers in Newton 1989 having different 3’-end were analogous to the degenerate primers using in Martin-Harris. Therefore, I disagree with Dr. Lennon’s conclusion that “Martin-Harris fails to disclose a plurality of different primers that have different hybridization sequences having different 3’ ends,”³⁰⁵ for the reasons described above.

³⁰² See Newton 1989at Abstract (emphasis added).

³⁰³ See *id.* at 2504 (emphasis added).

³⁰⁴ See *id.* at 2511 (emphases added).

³⁰⁵ See Lennon Rebuttal Report at ¶ 384.

b) [Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence]

159. As described above, I have reviewed the McSwiggen Report. Dr. Lennon also “reviewed the expert report of Dr. McSwiggen and agree with his methods, analysis and conclusion.”³⁰⁶

Under the permissive PCR conditions specified in Martin-Harris (i.e., annealing temperature equal to 35°C), and given the high level of degeneracy of the four types of primers (i.e., $4^5 = 1,024$ permutations), a POSA would have expected at least two different primers in each of the four pools of degenerate primers to anneal to the same sequence of template DNA.³⁰⁷

I disagree with the McSwiggen Report as it did not consider the Court’s construction of the term *target nucleic acid* in his analysis of the Martin-Harris primers.³⁰⁸ As described above, the Court’s construction for the term *target nucleic acid* is a “nucleic acid molecule of interest (e.g., a nucleic acid to be analyzed).” In Martin-Harris, the author’s strategy was to find local regions in the human genome that also contained APC, BRCA1 and myocillin genes.³⁰⁹ Based on their approach, Martin-Harris reported target nucleic acids in the range of 300 to 700 base pairs.³¹⁰ Dr. McSwiggen, however, considered the entire human genome,³¹¹ which is approximately 2,700 megabase pairs,³¹² which is roughly 3.8M times larger than the target nucleic acids of Martin-Harris, and most, if not all, of the hits would be expected to be off-target events. Thus, the analysis

³⁰⁶ See *id.* at ¶ 386.

³⁰⁷ See Lennon Rebuttal Report at ¶ 347 (emphasis added).

³⁰⁸ See McSwiggen Report at ¶¶ 9-10, purporting to rely on the Court’s construed terms “plurality of different primers,” “the same sequence,” “double-stranded target nucleic acid,” and the ’597 Patent definitions “anneal” and “complementary.”

³⁰⁹ See Martin-Harris at 309.

³¹⁰ See *id.* at Figure 2.

³¹¹ See McSwiggen Report at ¶ 64.

³¹² See Yu *et al.*, *Draft sequence of the rice genome (Oryza sativa L. ssp. Indica)*, SCIENCE 296:79-92 (2002) (“Yu”) at Abstract.

provided by Dr. McSwiggen and relied on by Dr. Lennon is flawed and should not be afforded any weight.

c) [Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence]

160. Dr. Lennon argues that “no portion of any F2 primer anneals to any portion of the complement of what Dr. Metzker identifies as the target-specific hybridization sequence of the F1 primer” based on an incorrect application of the Court’s claim construction for “target-specific hybridization sequence” to this limitation.³¹³ The limitation in question for Claim 1[e] is not “target-specific hybridization sequence” but rather the “complement of the target-specific hybridization sequence.” The Court did not construe “complement of the target-specific hybridization sequence” and Dr. Lennon has provided no basis for his argument that a primer that specifically anneals to the “complement of the target-specific hybridization sequence” must specifically bind to the portion of the complement that is complementary to the sequence of the “target-specific primer.” This narrow interpretation is inconsistent with his broad interpretations of other claim terms as discussed throughout this report and in my Opening Report. As discussed in paragraphs 562-566 of my Opening Report, Martin-Harris subjects the extension product to an amplification reaction using a primer that specifically anneals to the complement of the target-specific hybridization sequence (*i.e.*, the F2 primer). Fraiture also discloses this limitation for the reasons disclosed above in *supra* ¶ 146. Dr. Lennon has failed to point to any construction by the Court that is inconsistent with this disclosure. For at least the foregoing reasons, his analysis is flawed and should not be afforded any weight.

³¹³ See Lennon Rebuttal Report at ¶¶ 392-397.

d) Dependent Claims

161. Dr. Lennon argues that I failed “to establish that dependent claims 3-5, 10, 11, 13-27, and 29 are anticipated or obvious”³¹⁴ based on my his analysis of claim 1 of the ’597 Patent. As described above, Dr. Lennon’s assertions are incorrect, and Fraiture anticipates claim 1 of the ’597 Patent or renders it obvious in combination with the prior art. I also incorporate my reasons why Markham anticipates claims 3-5, 10, 11, 13-27, and 29 of the ’597 Patent or renders them obvious in combination with the prior art.³¹⁵

6. The APAGene Manual Anticipates the Asserted Claims or Renders Them Obvious in Combination with Other Prior Art

162. After reviewing the Lennon Rebuttal Report, my opinion has not changed that APAGene Manual anticipates the Asserted Claims of the ’597 Patent or renders them obvious in combination with other prior art references. Below, I address specific points raised by Dr. Lennon in his Rebuttal Report.

a) Claim 1[c]: Wherein the Different Hybridization Sequences Have Different 3’ Ends

163. Dr. Lennon contends that “[b]y contrast, the 3’ ends of the hybridization sequence of each DRT primer do not vary.”³¹⁶ I disagree as this is not the manner the prior art describes different 3’-ends. For example, Newton 1989 described an improved PCR method that utilized mismatched nucleotides on or near the 3’-end of primers resulting in different 3’-ends. Newton 1989 claimed that “[t]he basis of the invention is that unexpectedly, *oligonucleotides with a mismatched 3’-residue* will not function as primers in the PCR under appropriate conditions.”³¹⁷

³¹⁴ See Lennon Rebuttal Report ¶ 398.

³¹⁵ See Metzker Opening Report ¶¶ 569-635.

³¹⁶ See Lennon Rebuttal Report at ¶ 402.

³¹⁷ See Newton 1989at Abstract (emphasis added).

Newton 1989 observed that the mismatches did not have to occur directly at the 3'-end but can also occur internal to the 3'-end and still function as primers having different 3'-ends.

In some instances a single 3'-mismatched base does allow amplification to proceed. *We have shown that introducing additional deliberate mismatches near the 3' end of appropriate primers ameliorates this problem.*³¹⁸

Newton 1989 demonstrates internal mismatches up to seven bases from the 3'-end with improved specificity with primers having different 3'-ends.

In an attempt to increase the specificity of the ARMS primers *we chose to deliberately introduce an additional mismatch near their 3'-ends*. When primers 7f and 7g which have *a deliberate C/T mismatch seven bases from the 3'-end* were introduced to replace primers 7 and 7a (figure 2) specificity was improved.³¹⁹

164. One of ordinary skill in the art would have understood that the primers in Newton 1989 having different 3'-end were analogous to the degenerate primers using in Martin-Harris. Therefore, I disagree with Dr. Lennon's conclusion that "Fraiture and the APAGene Manual fail to disclose a plurality of different primers that have different hybridization sequences having different 3' ends,"³²⁰ for the reasons described above.

b) [Claim 1[c]: Wherein Each Primer of the Plurality of Different Primers Does Not Anneal to the Same Sequence]

165. As described above, I have reviewed the McSwiggen Report. Dr. Lennon also "reviewed the expert report of Dr. McSwiggen and agree with his methods, analysis and conclusion."³²¹

As Dr. McSwiggen explains, his analysis for the APAGene Manual is identical to that of Fraiture because the APAGene Manual's objective,

³¹⁸ See *id.* at 2504 (emphasis added).

³¹⁹ See *id.* at 2511 (emphases added).

³²⁰ See Lennon Rebuttal Report at ¶ 404.

³²¹ See *id.* at ¶ 406.

primer sets, and reaction conditions are the same, except that the APAGene Manual does not provide a gene target, species or experimental results.³²²

166. As Dr. McSwiggen explains his analysis of APAGene Manual is identical to that of Fraiture, I incorporate my comments regarding the Fraiture analysis, *see supra* at ¶ 145.

c) [Claim 1[e]: Step (c)(ii): A Primer That Specifically Anneals to the Complement of the Target-Specific Hybridization Sequence]

167. Dr. Lennon argues that “the GSPb primer does not anneal to the complement of the target-specific hybridization sequence of the GSPa primer” based on an incorrect application of the Court’s claim construction for “target-specific hybridization sequence” to this limitation.³²³ The limitation in question for Claim 1[e] is not “target-specific hybridization sequence” but rather the “complement of the target-specific hybridization sequence.” The Court did not construe “complement of the target-specific hybridization sequence” and Dr. Lennon has provided no basis for his argument that a primer that specifically anneals to the “complement of the target-specific hybridization sequence” must specifically bind to the portion of the complement that is complementary to the sequence of the “target-specific primer.” This narrow interpretation is inconsistent with his broad interpretations of other claim terms as discussed throughout this report and in my Opening Report. As discussed in paragraphs 650-652 of my Opening Report, APAGene Manual discloses this limitation. Dr. Lennon has failed to point to any construction by the Court that is inconsistent with this disclosure. For at least the foregoing reasons, his analysis is flawed and should not be afforded any weight.

³²² *See id.*

³²³ *See* Lennon Rebuttal Report at ¶¶ 408-412.

d) Dependent Claims

168. Dr. Lennon argues that I failed “to establish that dependent claims 3-5, 10, 11, 13-27, and 29 are anticipated or obvious”³²⁴ based on my his analysis of claim 1 of the ’597 Patent. As described above, Dr. Lennon’s assertions are incorrect, and Fraiture anticipates claim 1 of the ’597 Patent or renders it obvious in combination with the prior art. I also incorporate my reasons why Markham anticipates claims 3-5, 10, 11, 13-27, and 29 of the ’597 Patent or renders them obvious in combination with the prior art.³²⁵

7. ’810 References

169. Dr. Lennon is interpreting the specification of the ’810 and ’828 patent to support an reading of the ’810 claims where the second target specific primer in the accused product anneals to tail sequence added by the first target specific primer. While I have stated previously that I do not believe this reading is supported, if one accepts that it is, then this element is present in the ’810 patent, which predates the ’597 patent.

VII. XII. SECONDARY CONSIDERATIONS

170. After reviewing the Lennon Rebuttal Report, my opinion has not changed that the Asserted Claims of the ’810 and ’597 Patents are invalid as obvious. Below, I address certain points raised by Dr. Lennon. It should be understood, however, that other points not specifically addressed here should not be inferred as unspoken agreements with the opinions of Dr. Lennon.

171. Dr. Lennon has failed to establish a nexus between the claimed inventions and the purported commercial success, industry praise, long-felt need, and skepticism.

³²⁴ See *id.* at ¶ 413.

³²⁵ See Metzker Opening Report ¶¶ 656-731.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on December 23, 2020.



Michael Metzker, Ph.D.

EXHIBIT 7

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

ARCHERDEX, LLC and THE GENERAL)
HOSPITAL CORPORATION d/b/a)
MASSACHUSETTS GENERAL HOSPITAL,)
)

Plaintiffs,)
)

v.)

C.A. No.)
)

QIAGEN SCIENCES, LLC, QIAGEN)
LLC f/k/a QIAGEN, INC., QIAGEN)
BEVERLY, LLC f/k/a QIAGEN)
BEVERLY, INC., QIAGEN)
GAITHERSBURG, INC., QIAGEN)
GMBH, QIAGEN N.V. and JONATHAN)
ARNOLD,)
)

18-1019-MN

Defendants.)
)

REMOTE ORAL AND VIDEOTAPED DEPOSITION OF
DR. MICHAEL METZKER
10th day of February, 2021
VOLUME 2
Pages 146 - 274

1 Q. All right. I'd like to direct you to
2 the Newton prior art reference.

3 MR. REINES: And why don't we make
4 Newton the next exhibit in order, which I believe
5 is Exhibit 10.

6 (Exhibit 10 introduced.)

7 BY MR. REINES:

8 Q. Let me know when you have Newton.

9 A. Well, it just showed up and I'm
10 downloading it right now. I'll let you know when
11 it's opened. I have Newton open.

12 Q. All right. Now, what -- what was the
13 reason why you in -- in -- well, how do you -- how
14 did you find Newton, and what were you looking for?

15 A. Oh. I've been familiar with this Newton
16 paper for quite a while.

17 Q. And what is Newton about?

18 A. Newton is about -- he's developing a
19 method called "amplification refractory mutation
20 system." It is a way of doing allele-specific PCR
21 where the 3' ends of primers can be different to
22 selectively extend one variant in the genome over
23 another.

24 And this reference is important because
25 Newton describes that it's not necessary to have

1 only the 3' most nucleotide be variable, but you
2 can introduce variations upstream in the 5'
3 direction and still observe allele-specific PCR
4 results.

5 Q. Why is that important to your opinions
6 in this case?

7 A. Because one of the claim elements
8 referred to having different 3' ends, and I believe
9 I'm responding to Dr. Lennon's assertion that that
10 means it's only the most 3' end nucleotide, and I
11 disagree with that, and Newton supports my opinion
12 of why Dr. Lennon is incorrect.

13 Q. Okay. If you turn to 2504 of Newton.

14 A. (Complied.)

15 Q. And specifically, about three-quarters
16 of the way down the page, it states: In some
17 instances, a single 3' mismatch base does allow
18 amplification to proceed.

19 Do you see that?

20 A. I do.

21 Q. And do you agree with that?

22 A. I don't have any reason to disagree with
23 what Newton is saying.

24 Q. And so is this an instance that you
25 would call mispriming, which is a term you were

1 using yesterday?

2 A. I don't think I ever used the word
3 "mispriming." This is an instance where you might
4 have a mismatch on the most 3' terminal nucleotide,
5 and Newton is describing you can still have some
6 amplification. This is still an undesired
7 reaction.

8 And the point of Newton's paper is to
9 overcome this problem by designing mismatches
10 upstream of the most 3' nucleotide so that you do
11 not get extension for the incorrect primer.

12 Q. In terms of the -- the mismatches,
13 what -- what do you call that when a primer
14 hybridizes to a sequence that's a mismatch? Do you
15 have a term that you like to use for that?

16 A. I don't quite understand the question
17 you're asking.

18 Q. Yes. You understand that sometimes
19 primers will hybridize to a DNA sequence where
20 there's mismatch basis, for example, at the 3' end?

21 A. I mean, primers can hybridize to
22 sequences and have a mismatch at the 3' end.

23 Q. What do you call that phenomena?

24 A. I would call that phenomenon a primer
25 binding to an incorrect sequence.

1 Q. And in Newton -- or if you turn to the
2 figure on page 2508, let me know when you have
3 that.

4 A. Yes, I'm there.

5 Q. And you see those Primers 7 -- 7
6 through 7g?

7 A. Yes.

8 Q. And there, what does the underlying base
9 represent?

10 A. That -- let me read to make sure I
11 have -- I think I know what it is, but let me just
12 make sure I'm saying exactly what Newton is saying.
13 (Examined exhibit.)

14 Newton refers to them as deliberately
15 introduced mismatches.

16 Q. And in the Primers 7 through 7g, some of
17 those mismatched primers would attach, and then at
18 some point, they wouldn't; is that correct? Let me
19 ask again.

20 Some of the primers that are mis -- that
21 have multiple mismatches at 7 through 7g will prime
22 and some won't; is that correct?

23 MR. NARAVAGE: Objection, form.

24 A. I guess I don't understand what you mean
25 by "some will prime and some will not."

1 BY MR. REINES:

2 Q. Can you tell from Newton which of these
3 primers will hybridize, notwithstanding mismatches,
4 and which won't?

5 A. Well, not with what Newton is showing
6 here. The reason why he's deliberately introducing
7 mismatches, because there are different sequences
8 that are characterized by these different
9 mismatches.

10 The assay is designed so that one primer
11 binds to one sequence, and when it's correct, it
12 will amplify.

13 Q. And you -- you're saying you don't
14 know -- you can't identify a -- for example, let me
15 ask this: In 7b, there's an intentional mismatch
16 at the third base position from the 3' end,
17 correct?

18 A. Correct.

19 Q. And is -- do you -- can you tell from
20 Newton whether that one is able to successfully
21 hybridize and extend?

22 A. You -- I think you can. I think Newton
23 describes it.

24 For example, if you go to page 2511, at
25 the very bottom, four lines from the bottom, it

1 says: When Primers b and Primer and -- I'm sorry.
2 When Primer 7b and 7c replaced Primer 7 and 7a in
3 the system, the desired specificity was
4 observed ...

5 And then it goes on to talk about the
6 behavior of those two primers.

7 Q. Okay. And for -- so based on your
8 review of Newton, which of the primers in 7
9 through 7g were able to be hybridized and extend,
10 even though they have intentional mismatches?

11 A. You asked 7 through 7g --

12 Q. Yeah.

13 A. -- was the question?

14 I am sure it is here in the paper. I
15 will point that out. I pointed out 7b and 7c
16 already. (Examined exhibit.)

17 Q. So far, 7b and 7c.

18 Anything else? Any others?

19 A. Yeah -- I'm -- I'm looking -- for
20 example, at the same page at 2511, he says: When
21 primers 7d and 7e, which have a deliberate A/G
22 mismatch five bases from their 3'-ends were
23 introduced to replace Primer 7 and 7a, similar
24 results were obtained. And he points to (Figure 3,
25 lanes 23 to 28), to those of Primers 7f and 7g.

1 And then above that, he talks about 7f
2 and 7g, and he says: Unfortunately, the normal
3 Primer (7f), still generates a small amount of
4 product with the ZZ DNA, (Figure 3, lane 21), but
5 much reduced with respect to the yield with the
6 mutant primer on ZZ DNA with equivalent internal
7 control products.

8 So here, he's showing that he -- I would
9 view this as demonstrating how primer design works
10 with real experimental data to determine whether
11 primers are binding to their appropriate templates
12 or not.

13 Q. Are you surprised that primers that have
14 -- intentionally have mismatches will still
15 hybridize and extend?

16 A. I'm not surprised. It's an undesired
17 result, and it's, just as Newton is doing, doing
18 redesign to mitigate that undesired result.

19 Q. And in terms of -- so does Newton
20 propose using any of his primers with mismatches in
21 them?

22 A. Well, there -- that's not quite what
23 he's doing. I mean, he does say he deliberately
24 introduces mismatches, but the goal is actually to
25 look at two different templates, two different

1 sequences, that have either one allele or the
2 other, and to have the primers be able to
3 discriminate between one or the other with a PCR
4 amplification product.

5 Q. Right.

6 And -- and Newton states that that
7 specificity is enhanced by including a mismatch as
8 long as it's in the right place, correct?

9 A. Well, it's not -- it's not a mismatch
10 because there is a sequence that there's a perfect
11 match. The mismatch that he's referring to is just
12 basically a SNP, a single nucleotide polymorphism,
13 within the genome; and for an autosome, you'll have
14 two copies, and you may have one copy of the
15 wild-type sequence and one copy of the mutant
16 sequence. Those are different sequences.

17 And these primers are designed to bind
18 to one or the other, and because he's -- he's
19 deliberately included what he calls a "mismatch,"
20 but it's basically two primers that have either one
21 sequence or another.

22 Q. Besides the potential mismatch at the 3'
23 end with the SNP, he also proposes including a
24 mismatch, a couple of positions over, into towards
25 the 5' end, is that correct, to increase

1 specificity?

2 A. Yeah, I -- again, I -- I know that he
3 has said the underline is the mismatch.

4 But if you actually look at the
5 sequences 7 -- Primer 7 through 7g, those are
6 individual sequences. There is no mismatches that
7 are designed in the primers. The point is whether
8 one primer or the other can discriminate between a
9 polymorphism in the sequence.

10 Q. So you're -- so you're contending that
11 in Newton when he underlines a base and calls it a
12 "mismatch" that that's just the polymorph -- that's
13 the SNP? That's your reading of Newton?

14 A. Not necessarily.

15 Q. In fact, doesn't he talk about
16 deliberately introducing mismatches to help
17 destabilize the hybridization process for better
18 specificity?

19 A. Can you point me where you're reading?

20 Q. All I'm -- I mean, I'm -- I'm just
21 looking at the description of Figure 2, but I'm
22 not -- that's not reading. That's understanding, I
23 think.

24 The -- just to be clear. Newton is --
25 is a article that you've been long -- you've long

1 known about, correct?

2 A. I have known about it for a while.

3 Q. Yeah. And you considered it important
4 to introduce into this case, correct?

5 A. Important to point out that the most 3'
6 nucleotide doesn't have to be the only variable
7 nucleotide to have different 3' sequences.

8 Q. Okay. Now, in terms of what Newton
9 teaches, Newton teaches that there might be
10 occasions where the SNP difference might not create
11 enough of a destabilization to create the effect
12 that you want, and that you might need to include
13 an additional mismatch base in order to ensure that
14 you can detect the SNP, correct?

15 A. I just want to take a look at the
16 outstrike again.

17 Q. Sure.

18 A. (Examined exhibit.)

19 Correct. He is introducing a deliberate
20 mismatch, but there's also a different -- there's
21 also a polymorphic base as well. So the -- it's
22 not just a mismatch. He is also -- the 3' end
23 nucleotide is also different, and even though these
24 are mismatching, Primer, for example, b and
25 Primer -- sorry, Primer 7b and Primer 7c still bind

1 to different sequences.

2 Q. And your understanding of what he was
3 testing was where the intentionally mismatched base
4 should be in order to optimize your specificity
5 when you're testing for the SNP or polymorphism
6 that you're referring to at the 3' end, correct?

7 A. He -- he is varying the position of the
8 mismatch base, but, again, testing two different
9 sequences to show specificity of one allele over
10 the other.

11 Q. Okay. And so Newton is an example of a
12 situation where a researcher is intentionally
13 creating a mismatch where there's substantial
14 complementarity, correct?

15 MR. NARAVAGE: Objection, form.

16 A. Yeah, I disagree with that.

17 BY MR. REINES:

18 Q. All right. Will you agree that Newton
19 is proposing introduction of an intentional
20 mismatch to increase the sensitivity when testing
21 for a polymorphism at a different location?

22 A. I disagree with that, too.

23 Q. All right. Well, let's just go back
24 to -- the deliberately introduced mismatches, are
25 you still maintaining that those are the SNP in

1 Newton?

2 A. No. You -- you -- you pointed out --
3 you're -- you're correct. The mismatch is upstream
4 at the polymorphic nucleotide, which appears on the
5 most 3' end. So there's two --

6 (Simultaneous conversation.)

7 BY MR. REINES:

8 Q. I gotcha.

9 A. -- there's two primers for each mismatch
10 that he's testing, and -- and so he's just looking
11 at the position -- varying the positions to see if
12 he can increase specificity in his PCR products.

13 Q. And any -- do you have any opinion as to
14 which of the primers -- well, let me -- do you
15 understand that Newton is stating that it increases
16 sensitivity for his system for detecting
17 polymorphisms by intentionally including a mismatch
18 upstream towards the 5' end?

19 A. He is introducing a mismatch upstream of
20 the most 3' nucleotide, but in the Newton system,
21 it's one primer and one sequence.

22 Q. Okay. Well, let's just -- is it a
23 bottom line thing, or do you deny that Newton is
24 proposing that you intentionally include a mismatch
25 to improve your sensitivity for SNP detection?

1 A. I -- I think I just answered that
2 question.

3 Q. Well, you -- you deny that, right?

4 A. I -- no, I did not deny that. I -- I
5 said that Newton is intentionally or deliberately,
6 as he puts, introducing a mismatch into a primer
7 that will bind to a sequence in a PCR reaction.

8 Q. Okay. And in this -- and when he does
9 that, the -- the primer will bind to the sequence
10 even though there is an intentional mismatch,
11 right?

12 A. It does bind to the sequence only when
13 it also hybridizes with the correct nucleotide for
14 the SNP that they are trying to assay.

15 Q. Right.

16 And in that situation, the -- there's
17 substantial complementarity between the primer and
18 the -- I think what you called the "wrong sequence"
19 or something like that before?

20 MR. NARAVAGE: Objection, form.

21 A. I -- I disagree with that. So --

22 (Simultaneous conversation.)

23 BY MR. REINES:

24 Q. And -- oh.

25 A. -- his primer, having an intentional

1 mismatch, will only bind -- bind to one of the two
2 alleles and be extended.

3 Q. Okay. Let's take the one that it binds
4 to. Okay?

5 A. Yes.

6 Q. In the instance where, by design, his
7 primer hybridizes to a mismatched sequence, there's
8 substantial complementarity; otherwise, it
9 wouldn't -- it wouldn't anneal, correct?

10 A. Can you be more specific, what is
11 binding to what?

12 Q. Yes. You've got a primer and it's
13 binding and there's no mismatch at the polymorphic
14 site, but there is an intentional mismatch in the
15 primer sequence.

16 Would you -- would you acknowledge that
17 that's -- that there's substantial complementarity
18 if you're achieving -- achieving hybridization
19 between a primer and a mismatched sequence?

20 A. By design, he has tested out his
21 experiment to -- to correctly bind and be extended
22 by polymerase. He did that by design.

23 Q. Even when there was an -- and that
24 occurred even -- successfully even where there was
25 an intentional mismatch, correct?

1 A. I'm sorry. What -- what occurred? I --
2 I don't understand your -- your question.

3 Q. There's success -- in -- in Newton, he
4 shows examples of successful hybridization and
5 extension where a primer is matching -- wait a sec.
6 Let me start again.

7 Newton shows hybridization and extension
8 between a primer and an intentionally mismatched
9 sequence, correct?

10 A. And I wouldn't characterize it that way.
11 He intentionally mismatches the primer with one of
12 two different terminal nucleotides hybridized to
13 one of two different template sequences as a way of
14 doing allele specific PCR for SNP detection.

15 Q. Right.

16 And so we have to figure out why you're
17 saying "no" to my question.

18 I'm saying to you: In the instance
19 where the SNP is a match, in that instance, you're
20 achieving hybridization even though there's an
21 intentional mismatch elsewhere in the sequence,
22 correct?

23 A. I guess I said "no" because I'm -- I'm
24 just trying to be accurate that the mismatch is
25 deliberately designed in the primer. It's not

1 deliberately designed in the template. That's the
2 thing you're trying to --

3 (Simultaneous conversation.)

4 BY MR. REINES:

5 Q. Oh. All right. Okay.

6 A. -- and I think that's what you've been
7 saying, and so that's why I was disagreeing with
8 you.

9 Q. Yeah. I think when people go off and
10 read the -- the questions I've asked, I hope that
11 they would -- they would see that, but that's what
12 we have a record for.

13 In the instance of -- Newton shows
14 examples where there are primers which are
15 mismatched with a sequence, but yet, they hybridize
16 and extend, correct?

17 A. Can you repeat that question, please?

18 MR. REINES: Would the court reporter
19 please read that back.

20 (Record read.)

21 A. I disagree with that.

22 BY MR. REINES:

23 Q. What's your disagreement based on?

24 A. Well, Newton's not showing primers,
25 plurality "primers," binding to a sequence. This

1 is -- this is a very -- this is basic PCR where you
2 have two primers, a forward and a reverse primer,
3 and one of those two primers is the Newton primer
4 that has an intentional mismatch with one or two --
5 one or the other SNP nucleotides on the 3'.

6 So this is not primers; this is a primer
7 binding to a sequence.

8 Q. Okay.

9 MR. REINES: And -- and these -- these
10 responses seem evasive to me, but I'll plunge on.
11 I'm just putting that on the record.

12 BY MR. REINES:

13 Q. There are -- there is an example of a
14 primer mismatched with the sequence that
15 nevertheless hybridizes and extends in Newton,
16 correct?

17 A. I would agree with that. There is "a"
18 primer and "a" sequence, correct.

19 Q. Okay. And that primer is shown as 7b;
20 is that correct?

21 A. That is one of the two primers that are
22 used in two separate PCR assays.

23 Q. And another one of the primers is 7d,
24 correct?

25 A. Well, I think he -- he discusses 7d, e,

1 f, g not having the desired specificities that was
2 observed with Primer 7b and 7c.

3 Q. Okay. Well, let's stick to 7b, then.

4 In the case of 7b, the researcher
5 intentionally wants to have a mismatch between the
6 primer and the sequence in his hybridization and
7 extension reaction, correct?

8 A. They do that by design.

9 Q. Okay. And do you call that
10 "mispriming"?

11 A. I actually would not call that
12 mispriming. He actually is priming to the right
13 position. That is by design, and that is an assay
14 that improves specificity. So I would not call
15 that "mispriming" at all.

16 Q. And in that instance, would you
17 acknowledge that the Primer 7b has substantial
18 complementarity with the sequence --

19 A. Would you --

20 Q. -- within the language of the patent?

21 MR. NARAVAGE: Objection, form.

22 A. Sorry. Now I'm confused. Which --
23 which language of the patent are we referring to?

24 BY MR. REINES:

25 Q. With -- with language of the patent

1 that -- that defines "annealing" as complementarity
2 or substantial complementarity.

3 Let -- let's take it outside the patent
4 and then we'll go in the patent and we'll see if
5 there's a difference for you.

6 In the instance with Primer 7b where
7 there's hybridization and extension on a mismatched
8 primer and sequence, would you agree that there's
9 substantial complementarity?

10 A. I -- I don't know what "substantial"
11 means, but I can -- I can tell you of all the
12 nucleotides present--and it looks like it's at
13 least 20--there's one nucleotide that mismatches
14 with the template, by design.

15 Q. Okay. So the phrase "substantial
16 complementarity," that's not a phrase that you're
17 familiar with, correct?

18 A. Well, I --I've read it in the patent.

19 Q. No, I mean, in your -- let's talk about
20 you. That's not a term you would use in -- in an
21 article, "substantial complementarity," right?

22 A. I mean, I -- I don't know for a hundred
23 percent, but it's not something I would typically
24 write in a paper.

25 Q. And it's not something you've seen in

EXHIBIT 8

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

ARCHERDX, LLC and THE GENERAL	)	
HOSPITAL CORPORATION d/b/a	)	
MASSACHUSETTS GENERAL HOSPITAL,	)	
	)	
Plaintiffs,	)	
	)	
v.	)	C.A. No. 18-1019-MN
	)	
QIAGEN SCIENCES, LLC, QIAGEN LLC	)	
f/k/a QIAGEN, INC., QIAGEN BEVERLY,	)	
LLC f/k/a QIAGEN BEVERLY, INC.,	)	
QIAGEN GAITHERSBURG, LLC f/k/a	)	
QIAGEN GAITHERSBURG, INC., QIAGEN	)	
GMBH, QIAGEN N.V., and	)	
JONATHAN ARNOLD,	)	
	)	
Defendants.	)	

EXPERT REPORT OF JAMES A. MCSWIGGEN, PhD

TABLE OF CONTENTS

I.	Scope of Work	1
II.	Qualifications	2
III.	The '597 Annealing Limitation	3
IV.	Analysis	4
A.	Fraiture.....	4
1.	Overview	4
2.	PCR Conditions.....	7
3.	Analysis of Whole Genome	13
B.	APAgene Manual.....	20
C.	McCarty	20
1.	Overview	20
2.	PCR Conditions.....	21
3.	Analysis of Mu-TIR Primers binding to their Reverse Complement	24
D.	Martin-Harris	27
1.	Overview	27
2.	PCR Conditions.....	27
3.	Analysis of Whole Genome	30
E.	Morley.....	34
1.	Overview	34
2.	Analysis of Whole Genome	35

constant 5' portion. According to the APAGene Manual, however, it is not part of the UAP-N1 binding sequence that is designed to match the rest of the 5' constant sequence, as Dr. Metzker indicates by highlighting in yellow the overlapping region between UAP-N1 and the DRT primers.

DRT Primer	Sequences
DRT Primer A:	5'-GAACACGCGTCGTTTACCTCCXXXXGXXXXXTAGT-3'
DRT Primer B:	5'-GAACACGCGTCGTTTACCTCCXXXXXTAGT-3'
DRT Primer C:	5'-GAACACGCGTCGTTTACCTCCXXXXXTCAT-3'
DRT Primer D:	5'-GAACACGCGTCGTTTACCTCCXXXXGXXXXXCCTG-3'
UAP-N1:	5'- AGT CGG GAA GCA GTG GTA TCA ACG CAG AGT GGC CAT TAC GGC C GAA CAC GCG TCG TTT ACC T -3'
UAP-N2:	5'- C CUG GAA GCA GTG GTA TCA ACG-3'

Metzker Report, ¶ 318.

31. A POSA would have understood that the CC sequence can participate in binding to the DNA template. Taking the CC sequence into account provides slightly better binding strength for DRT Primer B and DRT Primer C, which shifts a small fraction of their sequences above the specified 42°C annealing temperature for the reaction. To account for both possibilities, I analyzed the DRT primers both with and without the central CC dinucleotide included as part of the variable 3' portion of the primer. The total number of permutations of degenerate primers in each set is depicted below.

Table 1			
Set	DRT Primer	Sequence	# Pooled Seqs
MixA	A	GAACACGCGTCGTTTACCT CC XXXXGXXXXXTAGT	262,144
MixB	B	GAACACGCGTCGTTTACCT CC XXXXXXTAGT	4,096
MixC	C	GAACACGCGTCGTTTACCT CC XXXXXXTCAT	4,096
MixD	D	GAACACGCGTCGTTTACCT CC XXXXXGXXXXXCCTG	1,048,576

target under the specified PCR conditions to confirm that at least some of the primer would bind to the target sequence.

41. Target sites having only one bound oligo under these conditions were removed from the resulting data set, leaving a total of 4,815 oligo/target pairs associated with 1,144 different targets. This demonstrates that some targets are bound by more than one oligo under these conditions. In fact, an average of 4.2 oligos are calculated to be bound by each site ($4816/1144$). This outcome is exactly what a POSA would have expected when using primers such as those disclosed in the APAGene Manual—and having a high degree of degeneracy—to anneal to templates under permissive annealing conditions (i.e., 42°C).

42. The difference in melting temperature was then determined between the interaction with the highest T_m and the T_m of other primers that annealed to the same sequence. Table 3 shows a small set of these primer pairs for each of the Primer DRT Mixes. Mixes A and D produce more instances in which different primers annealed to the same sequence. This was to be expected given that the average on-target T_m for Mixes A and D is well above the 42°C annealing temperature. By contrast, Mixes B and C have T_m values that are below the annealing temperature, for both exact matches and mismatches. Even under such conditions, however, different primers are predicted to anneal to the same sequence. Table 3 below demonstrates that each of the four DRT primer mixes contains primers that would bind to the same target sequence of the rice genome under the conditions specified in the APAGene Manual.

Table 3. Examples of multiple DRT primers binding to the same target sites

Mix	Description	Sequence	T _m	T _m Diff	% Bound
A	Motif	CCXXXXGXXXXTAGT			
	Template	GACGTGTGGGCCCGGTGGGATCTAGTATAATCCTCG			
	PairID 1	CCCGGgGGGATCTAGT	45.3	0	12.1
	PairID 2	CCCGGTGGGAaCTAGT	43.4	1.87	5.2
B	Motif	CCXXXXXXTAGT			
	Template	AAGACATGCTCCCCCGTAGTACTAGCATCG			
	PairID 4762	CCCCCgCGTAGT	35.3	0	0.37
	PairID 4763	CCCCgCCGTAGT	35.3	0	0.41
C	Motif	CCXXXXXXTCAT			
	Template	TTGCAAATTTCCGCGCATTTCATACAGAAAAGC			
	PairID 4788	CCGCGCgTTCAT	36.1	0	0.78
	PairID 4789	CCGCGCAgTCAT	35.6	0.5	0.61
D	Motif	XXXXXGXXXXCCTG			
	Template	CCATCTTCTTTTCGCGCAAGGCCGAGGAGAGCACA			
	PairID 1318	gaCGCGCAAGGCCtg	44.7	0	25.4
	PairID 1319	agCGCGCAAGGCCtg	44.4	0.3	22.3

43. A more complete list of primers that anneal to the same binding sequence of the rice genome and whose melting temperatures differ by less than 5°C is provided in the spreadsheet attached as Exhibit C. By limiting the number of primers examined and the number of target sequences in the rice genome in order to expedite the processing time, my analysis excludes many other instances in which more than one primer anneals to the same template sequence. Expanding the analysis to include all DRT Primers across the entire rice genome would have increased the number of such instances.

44. As demonstrated above, given the presence of primer pairs with T_m differences of 5°C or less and for which some percentage of each primer is predicted to bind to the target sequence, a POSA would have understood that none of the DRT Primer sequences in Fraiture and the APAGene Manual satisfy the requirement of '597 claim 1 that each primer of the

right column). McCarty does not list the sequences of those Mu-TIRs and neither does their preceding paper (Settles 2004 NAR 32_e54). Consequently, the DNA target for this analysis was chosen to be the reverse complement of the twelve degenerate primer sequences. It should be noted that McCarty sheared the maize DNA to about 1000 bp before starting the PCR analysis (McCarty Figure 1 and p8, right column).

54. **Primer Binding Sequence.** McCarty utilized a single degenerate pool of 12 primers having the following sequence: ACACGACGCTCTTCCGATCT NNNN CBCTCTTCKTCYATAATGGCAAT, where the sequence to the right of the NNNN is the Mu-specific sequence and where N=A,C,G,T; B= C,G,T; Y= C,T; and K= G,T. Page 9. The 12 Mu-specific sequences to the right of the NNNN were used for this analysis while the reverse complements of those sequences were used as the targets.

3. Analysis of Mu-TIR Primers binding to their Reverse Complement

55. I analyzed the affinity of Mu-TIR primer binding sites to their reverse complement sequence using OMP DE (DNA Software, Inc.), which uses a highly accurate nearest neighbor binding free energy model to determine binding affinity and T_m . Table 5 shows the predicted T_m for on-target binding of the Mu-TIR primers to their exact match reverse-complement target sequence. The T_m s range from a low of 67.1 °C for sequence 12 to a high of 75.4°C for sequence 5. These melting temperatures are 7-15°C above the higher of the two annealing temperatures used by McCarty (60°C) and 13-21°C above the lower of the two annealing temperatures (54°C). Under these conditions, the primers will be able to accommodate a significant number of mismatches in their interactions and still be mostly hybridized to the target sequence.

Table 5. Predicted binding of the 12 Mu-TIR primer sequences to their reverse complement sequence

SeqID	Query Seq	T _m
1	CCCTCTTCGTCCATAATGGCAAT	74.7
2	CCCTCTTCGTCTATAATGGCAAT	71.9
3	CCCTCTTCTTCCATAATGGCAAT	72.4
4	CCCTCTTCTTCTATAATGGCAAT	69.5
5	CGCTCTTCGTCCATAATGGCAAT	75.4
6	CGCTCTTCGTCTATAATGGCAAT	72.6
7	CGCTCTTCTTCCATAATGGCAAT	73.2
8	CGCTCTTCTTCTATAATGGCAAT	70.4
9	CTCTCTTCGTCCATAATGGCAAT	72.3
10	CTCTCTTCGTCTATAATGGCAAT	69.5
11	CTCTCTTCTTCCATAATGGCAAT	70.0
12	CTCTCTTCTTCTATAATGGCAAT	67.1
	Min	67.1
	Max	75.4

56. To evaluate Dr. Metzker's contention that the Mu-TIR primers satisfy the '597 Annealing Limitation, I analyzed whether any two Mu-TIR primers would anneal to the same reverse complement sequence of the primers listed in Table 5. The binding affinity and T_ms for each of the twelve Mu-TIR primer sequences was evaluated using ThermoBLAST (DNA Software, Inc.) against each of the twelve reverse complement sequences of those primers. A maximum of five interactions per primer sequence was returned. The results were further constrained by discarding interactions with T_ms below the annealing temperature of 60°C. That, in fact, did not remove any interactions because they all had T_ms above 60°C. Thus, on average, there were at least five primer sequences that can bind to **every** Mu-TIR target sequence with a T_m of 60°C or greater. Table 6 shows a sample of those results. The full list of primers that can anneal to each of the twelve Mu-TIR reverse complement sequences at T_ms in excess of 60°C can be found in Exhibit D.

Table 6. Examples of multiple Mu-TIR primers binding to same reverse complement sites

Target	Description	Sequence	T _m	T _m Diff	% Bound
2	Motif	CbCTCTTCKTCyATAATGGCAAT			
	Template	CCCTCTTCGTCTATAATGGCAAT			
	PairID 7	CCCTCTTCGTCTATAATGGCAAT	71.9	0	99.9
	PairID 8	CCCTCTTCGTCCATAATGGCAAT	66.8	5.0	89.6
9	Motif	CbCTCTTCKTCyATAATGGCAAT			
	Template	CTCTCTTCGTCCATAATGGCAAT			
	PairID 42	CTCTCTTCGTCCATAATGGCAAT	72.3	0	99.9
	PairID 43	CgCTCTTCGTCCATAATGGCAAT	70.0	2.3	99.4
10	Motif	CbCTCTTCKTCyATAATGGCAAT			
	Template	CTCTCTTCGTCTATAATGGCAAT			
	PairID 47	CTCTCTTCGTCTATAATGGCAAT	69.5	0	99.2
	PairID 48	CgCTCTTCGTCTATAATGGCAAT	67.0	2.5	95.4
11	Motif	CbCTCTTCKTCyATAATGGCAAT			
	Template	CTCTCTTCTTCCATAATGGCAAT			
	PairID 52	CTCTCTTCTTCCATAATGGCAAT	70.0	0	99.4
	PairID 53	CgCTCTTCTTCCATAATGGCAAT	67.6	2.5	96.4
12	Motif	CbCTCTTCKTCyATAATGGCAAT			
	Template	CTCTCTTCTTCTATAATGGCAAT			
	PairID 57	CTCTCTTCTTCTATAATGGCAAT	67.1	0	95.7
	PairID 58	CgCTCTTCTTCTATAATGGCAAT	64.5	2.6	77.8

57. As demonstrated above, given the presence of primer pairs with T_m differences of 5°C or less for which some percentage of each primer is predicted to bind to the target sequence, a POSA would have understood that none of the Mu-TIR primer sequences satisfy the requirement of '597 claim 1 that each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers.

69. Fifth, the amplification conditions in Martin-Harris are identical to those in Example 3 of Morley US20100248310. The methods state that “PCR was performed in a 25 μ L volume, with 50 ng total DNA...”.

70. For the foregoing reasons, a POSA would have understood that the target of these experiments is human genomic DNA. Martin-Harris does not specify that the DNA is sheared in any way, so the DNA template can be taken to mean entire chromosomes.

71. **Primer Binding Sequence.** Four different sets of primer sequences were used by Martin-Harris, as listed in Table 1, p309 of their paper. As shown in Table 7, below, the sequences have a constant 5' portion and a variable 3' portion. With 5 randomized positions (shown by N in the sequence below), there are 1024 possible sequences in the degenerate pool.

Table 7. Gene Walking Primer sequences from Martin-Harris, table 1

Set	Set Name	Sequence	# Pooled Seqs
GW-A	Gene Walking A	TGCTAGGATCCAAGG NNNNNATTCG	1,024
GW-T	Gene Walking T	TGCTAGGATCCAAGG NNNNNTTTCG	1,024
GW-C	Gene Walking C	TGCTAGGATCCAAGG NNNNNCTTCG	1,024
GW-G	Gene Walking G	TGCTAGGATCCAAGG NNNNNGTTCG	1,024

3. Analysis of Whole Genome

72. First, I searched the human genome for the four sets of GW sequences shown in Table 7. Table 8 shows the number of sequences found. The human genome contains a myriad of sites that exactly match the reverse complement of the primer sequences; they are tabulated in the column labeled “# Sites” in Table 8 below. Many of these binding sites are duplicated in the human genome, as is shown in the column labeled “# Duplicates” in Table 8 below. There are roughly 800 different exact match binding sites for **each** of the 1024 differences in the GW-A set. Essentially all of the GW-A sites in the human genome are duplicates of some other site in the genome with the exact same sequence.

Table 8. Martin-Harris Template-Binding Sites in the Human Genome (hg38)

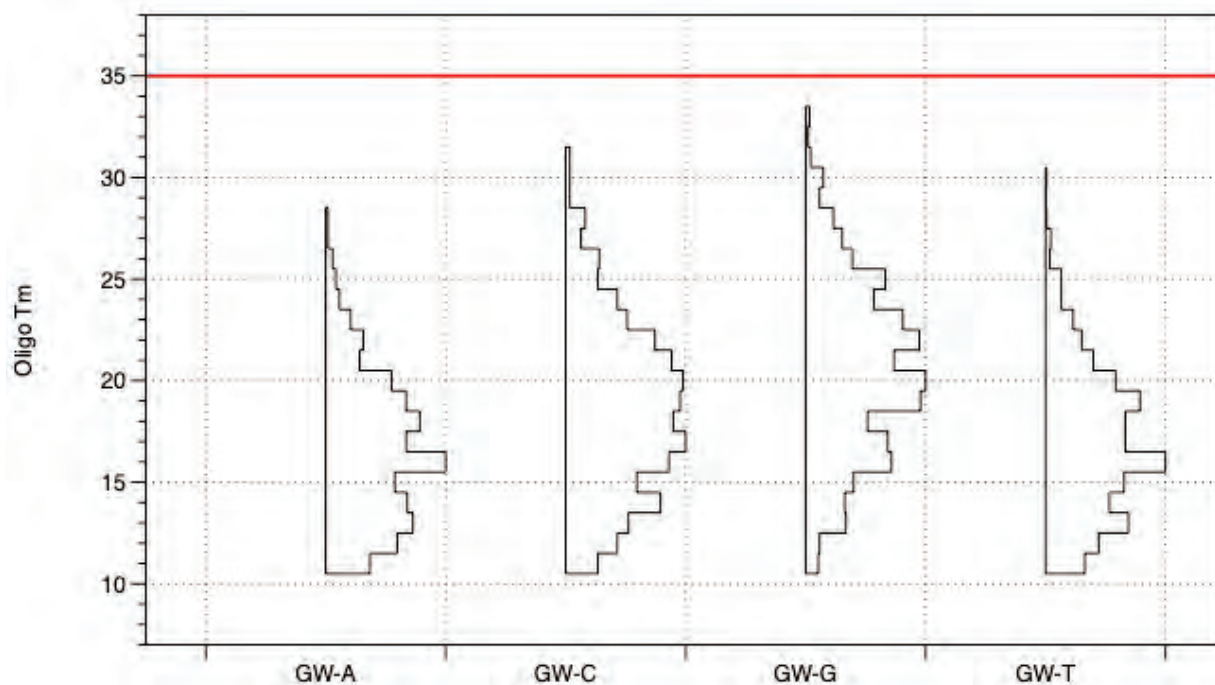
Set	# Sites	# Duplicates	# Pooled Seqs
GW-A	819,740	819,735	1,024
GW-T	1,149,515	1,149,510	1,024
GW-C	979,945	979,940	1,024
GW-G	684,168	681,468	1,024

73. As indicated in the '597 patent, a primer's predicted melting temperature (T_m) can be determined using any of a number of algorithms. 19:59-20:9. The '597 patent explains the relationship between binding specificity, predicted T_m and annealing temperature.

For most amplification regimens, the annealing temperature is selected to be about 5°C below the predicted T_m , although temperatures closer to and above the T_m (e.g., between 1°C and 5°C below the predicted T_m or between 1°C and 5°C above the predicted T_m) can be used, as can for example, temperatures more than 5°C below the predicted T_m (e.g., 6°C below, 8°C below, 10°C below or lower). In some embodiments, the closer an annealing temperature is to the T_m , the more specific is the annealing.

'597 patent, 20:20-29.

74. I analyzed the affinity of each of the GW primer binding sites against their reverse complement target sequence using OMP DE (DNA Software, Inc.), which uses a highly accurate nearest neighbor binding free energy model. Figure 3 below depicts the predicted T_m for on-target binding of the GW-A primers to their exact match reverse-complement target sequence with additional flanking G nucleotides. The target sequence included overhanging 5' and 3' G residues to mimic the unpaired flanking nucleotides that would be found in binding to larger DNA sequences. Flanking G and A residues provide the greatest amount of additional binding energy to duplexes, so the resulting T_m s will show the best-case conditions for these oligos. The plots show the probability distribution of melting temperatures along the Y axis for each of the GW primers. The red line indicates the 35°C annealing temperature used by Martin-Harris.

Figure 3. Predicted binding of the GW primer sequences to their reverse complement sequence

75. As depicted above, the range of melting temperatures is large, and significantly exceeds the 5°C difference between T_m and annealing temperature described in the '597 patent. The GW primers have predicted T_m s that span 17°C for GW-A up to 22°C for GW-G. In fact, the T_m range is likely actually even larger than observed here, but the full range is obscured due to truncation of the lower part of the distribution. As mentioned above, the OMP DE software has a low-end limit of about 10°C for estimating T_m s without first resetting the fitting parameters.

76. To evaluate Dr. Metzker's contention that the GW primers satisfy the '597 Annealing Limitation, I analyzed whether any two GW primers would anneal to the same sequence of the human genome. The binding affinity and T_m s for each set of the 1,024 GW primer sequences was evaluated using ThermoBLAST (DNA Software, Inc.) against human (hg38) genomic sequence. To make the analysis more manageable, I searched only those sections of the human genome that contained a perfect binding site to one of the GW primers, as enumerated in Table 8, above. That search represented 0.35-0.55% of the 3 GB human genome.

This was done to speed the analysis, knowing that the conditions specified by Martin-Harris virtually guarantee that multiple such instances will be found regardless.

77. A maximum of 200 interactions per primer sequence was returned. The results were further constrained by discarding interactions with T_m s below 25°C for sets GW-A and GW-T, and discarding interactions with T_m s below 30°C for sets GW-C and GW-G. These T_m s are just at the maximum of T_m s shown in Figure 3. Table 9 shows a sample of those results. Taking into account the parties' agreement that the phrase "the same sequence" is to be interpreted as "the identical sequence," these results show that multiple different primers from each of the Gene Walking sets are capable of binding to the same template sequences. The full list of 8,844 interactions targeting 3,181 different target sequences having two or more primers that can anneal at T_m s in excess of 25°C (GW-A or GW-T) and 30 °C (GW-C or GW-G) can be found in Exhibit E.

Table 9. Example of multiple GW primers binding to the same human genome sequence

Target	Description	Sequence	T_m	T_m Diff	% Bound
GW-A	Motif	nnnnnATTCG			
	Template	CTTGCCTCACTGCCATTGCTGCCCTGTC			
	PairID 8195	aGCCCATTTCG	28.6	0	0.40
	PairID8196	gGCCCATTTCG	28.1	0.54	0.33
GW-T	Motif	nnnnnTTTCG			
	Template	TATGCCTTTTTTCGCCTTTTCGTTT			
	PairID 8734	aCGCCTTTTCG	29.6	2.1	0.81
	PairID 8735	cCGCCTTTTCG	29.1	2.6	0.65
GW-C	Motif	nnnnnC TTCG			
	Template	CTCTCAAAAATGCCCTTCGTACCCCCACC			
	PairID 39	aGCCCC TTCG	32.8	1.22	2.86
	PairID 40	gGCCCC TTCG	32.3	1.74	2.34
GW-G	Motif	nnnnnG TTCG			
	Template	ATCTATTTTGCGCGGTTTCGTAGTATCTCT			
	PairID 6302	aGCGCG TTCG	38.3	3.81	15.5
	PairID 6303	tGCGCG TTCG	37.4	4.76	10.6

78. As demonstrated above, given the presence of primer pairs with T_m differences of 5°C or less for which some percentage of each primer is predicted to bind to the target sequence, a POSA would have understood that none of the GW primers satisfy the requirement of '597 claim 1 that each primer of the plurality of different primers does not anneal to the same sequence of the double-stranded target nucleic acid as any other primer of the plurality of different primers.

E. Morley

1. Overview

79. The Morley patent describes a “Method of DNA Amplification.” Abstract. The invention describes a method for amplifying a nucleic acid region using a PCR method designed to minimize the generation of amplicons from primers which have bound to nucleic acid regions other than the specific region of interest. The method is based on the determination that by rendering inefficient the functionality of either the forward primer or the reverse primer, the rate of amplification of irrelevant nucleic acid regions can be reduced relative to amplification of the region of interest. The approach was called Bottleneck PCR. Abstract.

80. Morley shows (Figure 11) that by rendering inefficient the functionality of either the forward primer or the reverse primer, the rate of amplification of irrelevant nucleic acid regions can be reduced relative to amplification of the region of interest.

81. Dr. Metzker contends that Figures 4 and 13 of Morley disclose the claimed plurality of different primers of '597 claim 1. Metzker Report, ¶¶ 459-461.

82. Figure 4 of Morley, below, shows either a degenerate primer-tag1 or a pool of primers binding to a site of the DNA template. Presumably, Dr. Metzker is not suggesting that the plurality of primers bind to the exact same site as suggested in Morley, Figure 4, since that

I declare under penalty of perjury that the foregoing is true and correct.

Dated: **November 20, 2020**

Signed: 
James McSwiggen, PhD

EXHIBIT 9

* HIGHLY CONFIDENTIAL *

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

ARCHERDX, INC., and THE GENERAL)
HOSPITAL CORPORATION, d/b/a)
MASSACHUSETTS GENERAL HOSPITAL,) C.A. No. 18-1019-MN
Plaintiffs,)
vs.)
QIAGEN SCIENCES, LLC, QIAGEN LLC)
f/k/a QIAGEN, INC., QIAGEN)
BEVERLY, LLC, f/k/a QIAGEN)
BEVERLY, INC., QIAGEN)
GAITHERSBURG, LLC, f/k/a QIAGEN)
GAITHERSBURG, INC., QIAGEN GMBH,)
QIAGEN N.V., and JONATHAN ARNOLD,)
Defendants.)

VIDEOCONFERENCED VIDEORECORDED DEPOSITION OF
JAMES McSWIGGEN, Ph.D.
January 22, 2021

HIGHLY CONFIDENTIAL

* HIGHLY CONFIDENTIAL *

1 MR. RUTT: Objection.

2 A. So imagine it this way: That -- the
3 Martin-Harris is trying to design this experiment and they
4 go and take these oligos to the synthesizer, people who
5 look at the sequences and say, Oh, do you realize that
6 this is only going to hit three genes? Do you think that
7 they would have ordered these oligos? No.

8 They want these things to be broadly
9 applicable to target all sorts of genes so that they can
10 apply it to any technology that they want to use it on.
11 It's the same thing here.

12 Q. (By Mr. Wu) And so you're saying that
13 despite that they are using gene-specific primers, that
14 every -- they can use the same primers for -- for any
15 target, or does it have to be adapted from experiment to
16 experiment?

17 MR. RUTT: Objection.

18 A. I'm saying the gene walking primers can be
19 used for any gene.

20 Q. (By Mr. Wu) But they are used with gene-
21 specific primers; is that correct?

22 A. In this experiment, yes. In this example,
23 yes.

24 Q. And in your report, when you were
25 discussing these primers that were binding to the same

Page 78

* HIGHLY CONFIDENTIAL *

1 sequence, did you identify primers binding to the sequence
2 of these three genes specifically?

3 A. I did identify primers binding to these
4 three chromosomes: Chromosome 1, 5, and 17. And those
5 examples are in the exhibit. I believe Exhibit E.

6 Q. Let's go to paragraph 84 of your report.
7 So go back to Exhibit 1.

8 A. 84. All right. I'm there.

9 Q. So in -- if you scroll to the -- the second
10 page, a long paragraph, in the middle of that paragraph,
11 you state, "Both Morley and Martin-Harris target the genes
12 APC, BRCA1, and myocillin and both fail to disclose the
13 species that is the subject of the analysis."

14 What do you mean when you say that Morley
15 and Martin-Harris target these genes?

16 A. That is -- that those ones were mentioned
17 in those documents.

18 Q. And what do you mean by "target"?

19 A. That those were some of the nucleic acids
20 of interest.

21 Q. Okay. I think, actually, we -- we can be
22 wrapping up early today, which I'm sure feels good.

23 Dr. McSwiggen -- Dr. McSwiggen, during the
24 day today, have you done everything you could to answer my
25 questions truthfully and to the best of your ability?

Page 79

INSTRUCTIONS FOR READING/CORRECTING YOUR DEPOSITION

To assist you in making corrections to your deposition testimony, please follow the directions below. If additional pages are necessary, please furnish them and attach the pages to the back of the errata sheet.

This is the final version of your deposition transcript.

Please read it carefully. If you find any errors or changes you wish to make, Insert the corrections on the errata sheet beside the page and line numbers.

If you are in possession of the original transcript, do NOT make any changes directly on the transcript.

Do NOT change any of the questions.

After completing your review, please sign the last page of the errata sheet, above the designated "Signature" line and return the transcript to your attorney.

ERRATA SHEET

Page Line

19 11 Change: "one structure of" to "unstructured"

Reason: I talked about a protein that binds to unstructured RNA

20 4 Change: "microscopy" to "spectrometry"

Reason: If I did say microscopy it was a mistake. I've never done fluorescence microscopy

21 20 Change: "check" to Cech

Reason: Name of my post-doctoral advisor

21 25 Change: "neutron" to "intron"

Reason:

23 11 Change: "sequencing" to "sequence"

Reason:

Page Line

23 12 Change: “plasma” to “plasmid”

Reason: _____

26 11 Change: “check” to “Cech”

Reason: _____

29 14 Change: “it's probably better” to “it's probably **no** better”

Reason: I believe I was saying the reverse of what is stated.

34 24 Change: “siRNA” to “Sirna”

Reason: Make it clear that I was talking about the company “Sirna”

63 19 Change: “con -- based pairs and the number of AT-based pairs” to
“**base pairs** and the number of AT **base pairs**”

Reason: don't know what the “con” would be, but I was talking about
base pairs

64 2 Change: “pre-energy” to “free energy”

Reason: _____

72 2 Change: “fusing” to “using”

XXX Subject to the above changes, I certify that the transcript is true and correct.

_____ No changes have been made. I certify that the transcript is true and correct.


(signature)

February 1, 2021
(date)

EXHIBIT 10



US010017810B2

(12) **United States Patent**
Iafrate et al.

(10) **Patent No.:** **US 10,017,810 B2**
(45) **Date of Patent:** ***Jul. 10, 2018**

(54) **METHODS FOR DETERMINING A NUCLEOTIDE SEQUENCE CONTIGUOUS TO A KNOWN TARGET NUCLEOTIDE SEQUENCE**

(71) Applicant: **THE GENERAL HOSPITAL CORPORATION**, Boston, MA (US)

(72) Inventors: **Anthony John Iafrate**, Newton, MA (US); **Long Phi Le**, Boston, MA (US); **Zongli Zheng**, Boston, MA (US)

(73) Assignee: **THE GENERAL HOSPITAL CORPORATION**, Boston, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **15/269,448**

(22) Filed: **Sep. 19, 2016**

(65) **Prior Publication Data**

US 2017/0226575 A1 Aug. 10, 2017

Related U.S. Application Data

(63) Continuation of application No. 13/793,564, filed on Mar. 11, 2013, now Pat. No. 9,487,828.

(60) Provisional application No. 61/645,364, filed on May 10, 2012, provisional application No. 61/679,302, filed on Aug. 3, 2012.

(51) **Int. Cl.**

C12Q 1/68 (2018.01)

C12Q 1/6855 (2018.01)

C12Q 1/6874 (2018.01)

C12Q 1/6806 (2018.01)

(52) **U.S. Cl.**

CPC **C12Q 1/6855** (2013.01); **C12Q 1/6806** (2013.01); **C12Q 1/6874** (2013.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,868,104 A 9/1989 Kurn et al.
6,087,101 A 7/2000 Gruelich et al.
6,172,214 B1 1/2001 Brenner
6,576,448 B2 6/2003 Weissman et al.
6,632,611 B2 10/2003 Su et al.
6,833,246 B2 12/2004 Balasubramanian
7,214,490 B2 5/2007 Su et al.
7,244,559 B2 7/2007 Rothberg et al.
7,273,730 B2 9/2007 Du Breuil
7,824,856 B2 11/2010 Montforte
9,487,828 B2* 11/2016 Iafrate C12Q 1/6874
2002/0086317 A1 7/2002 Nagayama
2003/0143553 A1 7/2003 Sommer
2009/0203085 A1 8/2009 Kurn et al.

2010/0286143 A1 11/2010 Dias-Santagata et al.
2012/0122701 A1 5/2012 Ryan et al.
2013/0005585 A1 1/2013 Anderson et al.
2013/0231253 A1 9/2013 Amorese et al.
2013/0303461 A1 11/2013 Iafrate et al.
2015/0011396 A1 1/2015 Schroeder et al.
2015/0140553 A1 5/2015 Cushing et al.

FOREIGN PATENT DOCUMENTS

AU 2005203617 A1 9/2005
WO 97/23646 A1 7/1997
WO 97/23647 A1 7/1997
WO 98/28443 A1 7/1998
WO 99/42618 A1 8/1999
WO 00/43544 A1 7/2000
WO 00/70095 A2 11/2000
WO 01/12859 A2 2/2001
WO 01/20035 A2 3/2001
WO 02/00938 A2 1/2002
WO 02/29117 A3 4/2002
WO 02/48402 A2 6/2002
WO 02/072772 A2 9/2002
WO 03/078645 A2 9/2003
WO 03/083435 A2 10/2003
WO 2004/011665 A2 2/2004
WO 2004/092418 A2 10/2004
WO 2005/065321 A2 7/2005
WO 2007/030759 A2 3/2007
WO 2007/057652 A1 5/2007
WO 2007/136717 A1 11/2007
WO 2008/005459 A2 1/2008
WO 2008/093098 A2 8/2008
WO 2009/102878 A2 8/2009
WO 2009/102896 A2 8/2009
WO 2009/117698 A2 9/2009
WO 2009/133466 A2 11/2009
WO 2009/148617 A2 12/2009
WO 2010/077288 A2 7/2010
WO 2010/083046 A2 7/2010

(Continued)

OTHER PUBLICATIONS

Gould et al., "Analysis of sequence variation among smeDEF multi drug efflux pump genes and flanking DNA from defined 16S rRNA subgroups of clinical *Stenotrophomonas maltophilia* isolates." *Journal of Antimicrobial chemotherapy* 54(2):348-353 (2004).
Liu et al., "High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences." *Biotechniques Rapid Dispatches* 43(5):649-656 (2007).
Wang et al., "Fusion primer and nested integrated PCR (FPNI-PCR): a new high-efficiency strategy for rapid chromosome walking or flanking sequence cloning." *BMC Biotechnology* 11(1):109 (2011).

(Continued)

Primary Examiner — Angela M. Bertagna

(74) *Attorney, Agent, or Firm* — Nixon Peabody LLP; David S. Resnick; Nicole D. Kling

(57) **ABSTRACT**

The technology described herein is directed to methods of determining oligonucleotide sequences, e.g. by enriching target sequences prior to sequencing the sequences.

20 Claims, 9 Drawing Sheets

US 10,017,810 B2

75

76

-continued

```

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 45
ccatctcctc cctgcgtgtc tcgattcag ctaaggtaac tggcgt      46

<210> SEQ ID NO 46
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 46
cgccaggttc cagtcacgac      20

<210> SEQ ID NO 47
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 47
aatgatacgg cgaccaccga gatctacact agatgcgaca ctctttccct acacgacgct      60
cttcgatct      70

<210> SEQ ID NO 48
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 48
gatcggaaga gccagttacc ttagctgagt cggagacacg cagggatgag atgg      54

```

What is claimed herein is:

1. A method for preparing a nucleic acid for sequencing, the method comprising:

- (i) ligating a universal oligonucleotide tail adaptor that comprises a first ligatable duplex end and a second unpaired end to a nucleic acid comprising a known target nucleotide sequence to produce a ligation product, the universal oligonucleotide tail adaptor comprising an amplification strand and a blocking strand, wherein a 3' duplex portion of the amplification strand and a 5' duplex portion of the blocking strand are substantially complementary and form the first ligatable duplex end;
- (ii) amplifying the ligation product using a first target-specific primer that specifically anneals to the known target nucleotide sequence and a first adaptor primer having a nucleotide sequence identical to a first portion of the amplification strand; and
- (iii) amplifying an amplification product of (ii) using a second target-specific primer that specifically anneals to the amplification product of (ii) and a second adaptor primer having a nucleotide sequence identical to a second portion of the amplification strand, wherein the second target-specific primer is nested relative to the first target-specific primer.

2. The method of claim 1, wherein the entirety of the blocking strand is a duplex portion that is substantially complementary to the 3' duplex portion of the amplification strand.

3. The method of claim 1, wherein the blocking strand further comprises a 3' unpaired portion which is not substantially complementary to the 5' unpaired portion of the amplification strand.

4. The method of claim 1, wherein the duplex portions of the amplification and blocking strands are at least 95% complementary, at least 98% complementary, at least 99% complementary, or 100% complementary.

5. The method of claim 1, wherein the nucleic acid is a deoxyribonucleic acid.

6. The method of claim 5, wherein the method further comprises subjecting a ribonucleic acid to a reverse transcriptase regimen to generate the deoxyribonucleic acid.

7. The method of claim 1 further comprising, prior to step (i), mechanically shearing a nucleic acid preparation to produce the nucleic acid comprising the known target nucleotide sequence.

8. The method of claim 7 further comprising, prior to step (i), end-repairing the nucleic acid to blunt the ends of the nucleic acid produced by mechanically shearing the nucleic acid preparation.

US 10,017,810 B2

77

9. The method of claim 7 further comprising, prior to step (i), phosphorylating the nucleic acid.

10. The method of claim 1 further comprising, prior to step (i), adenylating the nucleic acid to produce a 3'-adenosine overhang on the nucleic acid.

11. The method of claim 1, wherein ligating in step (i) comprises performing an overhang ligation reaction.

12. The method of claim 1, wherein ligating in step (i) comprises performing a TA ligation reaction.

13. The method of claim 1, wherein the first target-specific primer can specifically anneal to the known target nucleotide sequence at an annealing temperature in a range of 61° C. to 72° C.

14. The method of claim 1, wherein the known target nucleotide sequence comprises a sequence associated with a gene rearrangement.

15. The method of claim 1, wherein the universal oligonucleotide tail adaptor further comprises a barcode portion.

16. A method for preparing a nucleic acid for sequencing, the method comprising:

- (i) ligating a universal oligonucleotide tail adaptor that comprises a first ligatable duplex end and a second unpaired end to a nucleic acid comprising a known target nucleotide sequence to produce a ligation product, the universal oligonucleotide tail adaptor comprising an amplification strand and a blocking strand, wherein a 3' duplex portion of the amplification strand

78

and a 5' duplex portion of the blocking strand are substantially complementary and form the first ligatable duplex end;

- (ii) amplifying the ligation product using a first target-specific primer that specifically anneals to the known target nucleotide sequence and a first adaptor primer having a nucleotide sequence identical to a first portion of the amplification strand; and

- (iii) amplifying an amplification product of (ii) using a second target-specific primer that specifically anneals to the amplification product of (ii) and a second adaptor primer having a nucleotide sequence identical to a second portion of the amplification strand, wherein ligating in step (i) comprises performing an overhang ligation reaction, and wherein the universal oligonucleotide tail adaptor further comprises a barcode portion.

17. The method of claim 16 further comprising, prior to step (i), phosphorylating the nucleic acid.

18. The method of claim 16 further comprising, prior to step (i), adenylating the nucleic acid to produce a 3'-adenosine overhang on the nucleic acid.

19. The method of claim 16, wherein ligating in step (i) comprises performing a TA ligation reaction.

20. The method of claim 16 further comprising sequencing an amplification product of (iii) using a first sequencing primer and a second sequencing primer.

* * * * *

EXHIBIT 11



US009487828B2

(12) **United States Patent**
Iafrate et al.

(10) **Patent No.:** **US 9,487,828 B2**
(45) **Date of Patent:** **Nov. 8, 2016**

(54) **METHODS FOR DETERMINING A NUCLEOTIDE SEQUENCE CONTIGUOUS TO A KNOWN TARGET NUCLEOTIDE SEQUENCE**

(71) Applicant: **THE GENERAL HOSPITAL CORPORATION**, Boston, MA (US)

(72) Inventors: **Anthony John Iafrate**, Newton, MA (US); **Long Phi Le**, Boston, MA (US); **Zongli Zheng**, Boston, MA (US)

(73) Assignee: **The General Hospital Corporation**, Boston, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 5 days.

(21) Appl. No.: **13/793,564**

(22) Filed: **Mar. 11, 2013**

(65) **Prior Publication Data**

US 2013/0303461 A1 Nov. 14, 2013

Related U.S. Application Data

(60) Provisional application No. 61/645,364, filed on May 10, 2012, provisional application No. 61/679,302, filed on Aug. 3, 2012.

(51) **Int. Cl.**
C12Q 1/68 (2006.01)

(52) **U.S. Cl.**
CPC **C12Q 1/6874** (2013.01); **C12Q 1/6806** (2013.01); **C12Q 1/6855** (2013.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,868,104 A 9/1989 Kurn et al.
6,087,101 A 7/2000 Grulich et al.
6,172,214 B1 1/2001 Brenner
6,576,448 B2 6/2003 Weissman et al.
6,632,611 B2 10/2003 Su et al.
6,833,246 B2 12/2004 Balasubramanian
7,214,490 B2 5/2007 Su et al.
7,244,559 B2 7/2007 Rothberg et al.
7,273,730 B2 9/2007 Du Breuil Lastrucci
7,824,856 B2 11/2010 Monforte
2002/0086317 A1 7/2002 Nagayama
2003/0143553 A1 7/2003 Sommer
2010/0286143 A1 11/2010 Dias-Santagata et al.
2012/0122701 A1* 5/2012 Ryan et al. 506/2
2013/0231253 A1 9/2013 Amorese et al.
2013/0303461 A1 11/2013 Iafrate et al.
2015/0011396 A1 1/2015 Schroeder et al.
2015/0140553 A1 5/2015 Cushing et al.

FOREIGN PATENT DOCUMENTS

WO 9723646 A1 7/1997
WO 9723647 A1 7/1997

WO 9828443 A1 7/1998
WO 9942618 A1 8/1999
WO 0043544 A1 7/2000
WO 0070095 A2 11/2000
WO 0112859 A2 2/2001
WO 0120035 A2 3/2001
WO 0200938 A2 1/2002
WO 0229117 A2 4/2002
WO 0248402 A2 6/2002
WO 02072772 A2 9/2002
WO 03078645 A2 9/2003
WO 03083435 A2 10/2003
WO 2004011665 A2 2/2004
WO 2004092418 A2 10/2004
WO 2005065321 A2 7/2005
WO 2007030759 A2 3/2007
WO 2007/057652 A1 5/2007
WO 2007136717 A1 11/2007
WO 2008005459 A2 1/2008
WO 2008/093098 A2 8/2008
WO 2009102878 A2 8/2009
WO 2009102896 A2 8/2009
WO 2009117698 A2 9/2009
WO 2009/133466 A2 11/2009
WO 2009148617 A2 12/2009
WO 2010077288 A2 7/2010
WO 2010083046 A2 7/2010
WO 2011019964 A1 2/2011
WO 2011032053 A1 3/2011
WO 2011053987 A1 5/2011
WO 2011156529 A2 12/2011
WO 2012003374 A2 1/2012
WO 2012040387 A1 3/2012
WO 2012/044956 A1 4/2012
WO 2012103154 A1 8/2012
WO 2013059746 A1 4/2013
WO 2013074833 A1 5/2013
WO 2013112923 A1 8/2013
WO 2013191775 A2 12/2013
WO 2014144092 A1 9/2014
WO 2014150931 A1 9/2014
WO 2015073711 A1 5/2015

OTHER PUBLICATIONS

Zheng et al. *Nature Protocols* (2011) 6(9): 1367-1376.*
Yuanxin et al. *Nucleic Acids Research* (2003) 31(12): e68.*
Lamant et al. *Blood* (1999) 93(9): 3088-3095.*
Ali et al., *Infect. Immun.*, 68(12):6857-6864 (2000). "Sequence Analysis of TnpA Insertion Sites in *Vibrio cholerae* Mutants Defective in Rngose Polysaccharide Production."
Baetens et al., *Hum Mutat.*, 32:1053-1062 with Supplemental Data (2011). "Applying Massive Parallel Sequencing to Molecular Diagnosis of Marfan and Loeys-Dietz Syndromes."
Grace et al., *Analytical Biochemistry*, 263:85-92 (1998). "Degradable dUMP Outer Primers in Merged Tandem (M/T)-Nested PCR: Low- and Single-Copy DNA Target Amplification."

(Continued)

Primary Examiner—Angela M Bertagna

(74) *Attorney, Agent, or Firm*—Nixon Peabody LLP

(57) **ABSTRACT**

The technology described herein is directed to methods of determining oligonucleotide sequences, e.g. by enriching target sequences prior to sequencing the sequences.

47 Claims, 10 Drawing Sheets

EXHIBIT 12



US 20130303461A1

(19) **United States**(12) **Patent Application Publication**
Iafrate et al.(10) **Pub. No.: US 2013/0303461 A1**(43) **Pub. Date: Nov. 14, 2013**(54) **METHODS FOR DETERMINING A
NUCLEOTIDE SEQUENCE****Publication Classification**(71) Applicant: **THE GENERAL HOSPITAL
CORPORATION**, Boston, MA (US)(51) **Int. Cl.**
C12Q 1/68 (2006.01)(72) Inventors: **Anthony John Iafrate**, Newton, MA
(US); **Long Phi Le**, Boston, MA (US);
Zongli Zheng, Boston, MA (US)(52) **U.S. Cl.**
CPC **C12Q 1/6874** (2013.01)
USPC **514/19.3**; 506/2; 514/318; 514/86;
514/252.14; 514/232.8; 514/183; 514/245;
514/252.02; 514/265.1; 514/414; 514/346;
514/266.22; 514/312; 514/252.18; 514/275;
514/338; 514/333; 514/234.5; 514/172(73) Assignee: **THE GENERAL HOSPITAL
CORPORATION**, Boston, MA (US)(21) Appl. No.: **13/793,564**(22) Filed: **Mar. 11, 2013**(57) **ABSTRACT****Related U.S. Application Data**(60) Provisional application No. 61/645,364, filed on May
10, 2012, provisional application No. 61/679,302,
filed on Aug. 3, 2012.The technology described herein is directed to methods of
determining oligonucleotide sequences, e.g. by enriching tar-
get sequences prior to sequencing the sequences.

EXHIBIT 13

Manual of PATENT EXAMINING PROCEDURE

Original Ninth Edition, March 2014

Latest Revision June 2020 [R-10.2019]

U.S. DEPARTMENT OF
COMMERCE

United States Patent and
Trademark Office



Chapter		Page
	Change Summary	-1
	Title Page	-1
	Foreword	-1
	Introduction	-1
100	Secrecy, Access, National Security, and Foreign Filing	100-1
200	Types and Status of Application; Benefit and Priority Claims	200-1
300	Ownership and Assignment	300-1
400	Representative of Applicant or Owner	400-1
500	Receipt and Handling of Mail and Papers	500-1
600	Parts, Form, and Content of Application	600-1
700	Examination of Applications	700-1
800	Restriction in Applications Filed Under 35 U.S.C. 111; Double Patenting	800-1
900	Prior Art, Classification, and Search	900-1
1000	Matters Decided by Various U.S. Patent and Trademark Office Officials	1000-1
1100	Statutory Invention Registration (SIR); Pre-Grant Publication (PGPub) and Preissuance Submissions	1100-1
1200	Appeal	1200-1
1300	Allowance and Issue	1300-1
1400	Correction of Patents	1400-1
1500	Design Patents	1500-1
1600	Plant Patents	1600-1
1700	Miscellaneous	1700-1
1800	Patent Cooperation Treaty	1800-1
1900	Protest	1900-1
2000	Duty of Disclosure	2000-1
2100	Patentability	2100-1
2200	Citation of Prior Art and Ex Parte Reexamination of Patents	2200-1
2300	Interference and Derivation Proceedings	2300-1
2400	Biotechnology	2400-1
2500	Maintenance Fees	2500-1
2600	Optional Inter Partes Reexamination	2600-1
2700	Patent Terms, Adjustments, and Extensions	2700-1
2800	Supplemental Examination	2800-1
2900	International Design Applications	2900-1
FPC	Form Paragraphs Consolidated	FPC-1
I	Appendix I: [Reserved]	I-1
II	Appendix II: List of Decisions Cited	II-1
L	Appendix L: Patent Laws	L-1
R	Appendix R: Patent Rules	R-1
T	Appendix T: Patent Cooperation Treaty and Regulations Under the PCT	T-1
AI	Appendix AI: Administrative Instructions Under the PCT	AI-1
P	Appendix P: Paris Convention	P-1
Index	Subject Matter Index	Index-1

Chapter 700 Examination of Applications

701	Statutory Authority for Examination	706.02(c)	[Reserved]
702	Requisites of the Application	706.02(c)(1)	[Reserved]
702.01	Obviously Informal Cases		
703	[Reserved]	706.02(c)(2)	[Reserved]
704	Search and Requirements for Information	706.02(d)	[Reserved]
		706.02(e)	[Reserved]
704.01	Search	706.02(f)	[Reserved]
704.02	[Reserved]	706.02(f)(1)	[Reserved]
-704.09		706.02(f)(2)	[Reserved]
704.10	Requirements for Information	706.02(g)	[Reserved]
704.11	What Information May Be Required	706.02(h)	[Reserved]
704.11(a)	Examples of Information	706.02(i)	[Reserved]
	Reasonably Required	706.02(j)	[Reserved]
704.11(b)	When May a Requirement for Information Be Made	706.02(k)	[Reserved]
		706.02(l)	[Reserved]
704.12	Replies to a Requirement for Information	706.02(l)(1)	[Reserved]
		706.02(l)(2)	[Reserved]
704.12(a)	Relationship of Requirement for Information to Duty of Disclosure	706.02(l)(3)	[Reserved]
704.12(b)	What Constitutes a Complete Reply	706.02(m)	[Reserved]
		706.02(n)	[Reserved]
704.12(c)	Treatment of an Incomplete Reply	706.03	Rejections Not Based on Prior Art
		706.03(a)	[Reserved]
		706.03(b)	[Reserved]
704.13	Time Periods for Reply	706.03(c)	[Reserved]
704.14	Making a Requirement for Information	706.03(d)	[Reserved]
		706.03(e)	[Reserved]
704.14(a)	Format of the Requirement	706.03(f)	[Reserved]
704.14(b)	Examiner's Obligation Following Applicant's Reply	-706.03(j)	
		706.03(k)	[Reserved]
704.14(c)	Petitions to Requirements Under 37 CFR 1.105	706.03(l)	[Reserved]
		-706.03(n)	
704.14(d)	Relationship to Information Disclosure Statements	706.03(o)	[Reserved]
		706.03(p)	[Reserved]
705	Patentability Reports	-706.03(r)	
705.01	Instructions re Patentability Reports	706.03(s)	[Reserved]
705.01(a)	Nature of P.R., Its Use and Disposal	706.03(t)	[Reserved]
		706.03(u)	[Reserved]
705.01(b)	Sequence of Examination	706.03(v)	[Reserved]
705.01(c)	Counting and Recording P.R.s	706.03(w)	[Reserved]
705.01(d)	[Reserved]	706.03(x)	[Reserved]
705.01(e)	Limitation as to Use	706.03(y)	[Reserved]
705.01(f)	Interviews With Applicants	706.04	Rejection of Previously Allowed Claims
706	Rejection of Claims		
706.01	Contrasted With Objections	706.05	Rejection After Allowance of Application
706.02	Rejection on Prior Art		
706.02(a)	[Reserved]	706.06	Rejection of Claims Copied From Patent
706.02(a)(1)	[Reserved]		
706.02(a)(2)	[Reserved]	706.07	Final Rejection
706.02(b)	[Reserved]	706.07(a)	Final Rejection, When Proper on Second Action
706.02(b)(1)	[Reserved]		
706.02(b)(2)	[Reserved]		

706.07(b)	Final Rejection, When Proper on First Action	707.09	Signing by Primary or Other Authorized Examiner
706.07(c)	Final Rejection, Premature	707.10	Entry
706.07(d)	Final Rejection, Withdrawal of, Premature	707.11	Date
706.07(e)	Withdrawal of Final Rejection, General	707.12	Mailing
706.07(f)	Time for Reply to Final Rejection	707.13	Returned Office Action
706.07(g)	Transitional After-Final Practice	708	Order of Examination
706.07(h)	Request for Continued Examination (RCE) Practice	708.01	List of Special Cases
707	Examiner's Letter or Action	708.02	Petition To Make Special
707.01	Primary Examiner Indicates Action for New Assistant	708.02(a)	Accelerated Examination
707.02	Applications Up for Third Action and 5-Year Applications	708.02(b)	Prioritized Examination
707.03-707.04	[Reserved]	708.02(c)	Patent Prosecution Highway Program
707.05	Citation of References	708.03	Examiner Tenders Resignation
707.05(a)	Copies of Cited References	709	Suspension of Action
707.05(b)	Citation of Related Art and Information by Applicants	709.01	Overlapping Applications by Same Applicant or Owned by Same Assignee
707.05(c)	Order of Listing	710	Period for Reply
707.05(d)	Reference Cited in Subsequent Actions	710.01	Statutory Period
707.05(e)	Data Used in Citing References	710.01(a)	Statutory Period, How Computed
707.05(f)	[Reserved]	710.02	Shortened Statutory Period and Time Limit Actions Computed
707.05(g)	Incorrect Citation of References	710.02(a)	[Reserved]
707.06	Citation of Decisions, Orders Memorandums, and Notices	710.02(b)	Shortened Statutory Period: Situations in Which Used
707.07	Completeness and Clarity of Examiner's Action	710.02(c)	Specified Time Limits: Situations in Which Used
707.07(a)	Complete Action on Formal Matters	710.02(d)	Difference Between Shortened Statutory Periods for Reply and Specified Time Limits
707.07(b)	[Reserved]	710.02(e)	Extension of Time
707.07(c)		710.03	[Reserved]
707.07(d)	Language To Be Used in Rejecting Claims	710.04	Two Periods Running
707.07(e)	Note All Outstanding Requirements	710.04(a)	Copying Patent Claims
707.07(f)	Answer All Material Traversed	710.05	Period Ending on Saturday, Sunday, or a Federal Holiday
707.07(g)	Piecemeal Examination	710.06	Situations When Reply Period Is Reset or Restarted
707.07(h)	Notify of Inaccuracies in Amendment	711	Abandonment of Patent Application
707.07(i)	Each Claim To Be Mentioned in Each Office Action	711.01	Express or Formal Abandonment
707.07(j)	State When Claims Are Allowable	711.02	Failure To Take Required Action During Statutory Period
707.07(k)	Numbering Paragraphs	711.02(a)	Insufficiency of Reply
707.07(l)	Comment on Examples	711.02(b)	Special Situations Involving Abandonment
707.08	Reviewing and Initialing by Assistant Examiner	711.02(c)	Termination of Proceedings
		711.03	Reconsideration of Holding of Abandonment; Revival
		711.03(a)	Holding Based on Insufficiency of Reply

711.03(b)	Holding Based on Failure To Reply Within Period	714.02	Must Be Fully Responsive
711.03(c)	Petitions Relating to Abandonment	714.03	Amendments Not Fully Responsive, Action To Be Taken
711.03(d)	Examiner's Statement on Petition To Set Aside Examiner's Holding	714.03(a)	Supplemental Amendment
711.04	Public Access to Abandoned Applications	714.04	Claims Presented in Amendment With No Attempt To Point Out Patentable Novelty
711.04(a)	Date of Abandonment	714.05	Examiner Should Immediately Review
711.04(b)	Ordering of Patented and Abandoned Files	714.06	[Reserved]
711.04(c)	Notifying Applicants of Abandonment	714.07	Amendments Not in Permanent Ink
711.05	Letter of Abandonment Received After Application Is Allowed	714.08	[Reserved]
711.06	Abstracts, Abbreviations, and Defensive Publications	-714.09	
711.06(a)	Citation and Use of Abstracts, Abbreviations, and Defensive Publications as References	714.10	Claims Added in Excess of Claims Previously Paid For
712	[Reserved]	714.11	Amendment Filed During Interference Proceedings
713	Interviews	714.12	Amendments and Other Replies After Final Rejection or Action
713.01	General Policy, How Conducted	714.13	Amendments and Other Replies After Final Rejection or Action, Procedure Followed
713.02	Interviews Prior to First Official Action	714.14	Amendments After Allowance of All Claims
713.03	Interview for "Sounding Out" Examiner Not Permitted	714.15	Amendment Received in Technology Center After Mailing of Notice of Allowance
713.04	Substance of Interview Must Be Made of Record	714.16	Amendment After Notice of Allowance, 37 CFR 1.312
713.05	Interviews Prohibited or Granted, Special Situations	714.16(a)	Amendments Under 37 CFR 1.312, Copied Patent Claims
713.06	No Inter Partes Questions Discussed Ex Parte	714.16(b)	Amendments Under 37 CFR 1.312 Filed With a Motion Under 37 CFR 41.208
713.07	Exposure of Other Cases	714.16(c)	Amendments Under 37 CFR 1.312, Additional Claims
713.08	Demonstration, Exhibits, Models	714.16(d)	Amendments Under 37 CFR 1.312, Handling
713.09	Interviews Between Final Rejection and Notice of Appeal	714.16(e)	Amendments Under 37 CFR 1.312, Entry in Part
713.10	Interview Preceding Filing Amendment Under 37 CFR 1.312		
714	Amendments, Applicant's Action		
714.01	Signatures to Amendments	714.17	Amendment Filed After the Period for Reply Has Expired
714.01(a)	Unsigned or Improperly Signed Amendment	714.18	Entry of Amendments
714.01(b)	[Reserved]	714.19	List of Amendments, Entry Denied
714.01(c)	Signed by Attorney or Agent Not of Record	714.20	List of Amendments Entered in Part
714.01(d)	Amendment Signed by Applicant but Not by Attorney or Agent of Record	714.21	Amendments Inadvertently Entered, No Legal Effect
714.01(e)	Amendments Before First Office Action	714.22	[Reserved]
		-714.24	
		714.25	Discourtesy of Applicant or Attorney

715	Swearing Behind a Reference — Affidavit or Declaration Under 37 CFR 1.131(a)	716.01(d)	Weighing Objective Evidence
		716.02	Allegations of Unexpected Results
		716.02(a)	Evidence Must Show Unexpected Results
715.01	37 CFR 1.131(a) Affidavits Versus 37 CFR 1.132 Affidavits	716.02(b)	Burden on Applicant
715.01(a)	Reference Is a Patent or Published Application Naming Different Inventive Entity With at Least One Common Inventor	716.02(c)	Weighing Evidence of Expected and Unexpected Results
		716.02(d)	Unexpected Results
715.01(b)	Reference and Application Have Common Assignee	716.02(e)	Commensurate in Scope With Claimed Invention
715.01(c)	Reference Is Publication of Applicant's Own Invention	716.02(f)	Comparison With Closest Prior Art
715.01(d)	Activities Applied Against the Claims	716.02(g)	Advantages Disclosed or Inherent Declaration or Affidavit Form
		716.03	Commercial Success
715.02	How Much of the Claimed Invention Must Be Shown, Including the General Rule as to Generic Claims	716.03(a)	Commercial Success
			Commensurate in Scope With Claimed Invention
715.03	Genus-Species, Practice Relative to Cases Where Predictability Is in Question	716.03(b)	Commercial Success Derived From Claimed Invention
		716.04	Long-Felt Need and Failure of Others
715.04	Who May Make Affidavit or Declaration Under 37 CFR 1.131(a); Formal Requirements of Affidavits and Declarations	716.05	Skepticism of Experts
		716.06	Copying
		716.07	Inoperability of References
715.05	U.S. Patent or Application Publication Claiming Same Invention	716.08	Utility and Operability of Applicant's Disclosure
715.06	[Reserved]	716.09	Sufficiency of Disclosure
715.07	Facts and Documentary Evidence	716.10	Attribution Affidavit or Declaration to Overcome Rejection Under Pre-AIA 35 U.S.C. 102 or 103
715.07(a)	Diligence		
715.07(b)	Interference Testimony Sometimes Used	717	Prior Art Exceptions under AIA 35 U.S.C. 102(b)(1) and (2)
715.07(c)	Acts Relied Upon Must Have Been Carried Out in This Country or a NAFTA or WTO Member Country	717.01	Affidavit or Declaration Under 37 CFR 1.130
		717.01(a)	Declarations or Affidavits under 37 CFR 1.130(a) – Attribution
715.07(d)	Disposition of Exhibits	717.01(a)(1)	Evaluation of Declarations or Affidavits under 37 CFR 1.130(a)
715.08	Decided by Primary Examiner		
715.09	Timely Presentation		
715.10	Review of Affidavit or Declaration for Evidence of Prior Public Use or Sale or Failure to Disclose Best Mode	717.01(b)	Declarations or Affidavits under 37 CFR 1.130(b) – Prior Public Disclosure
716	Affidavits or Declarations Traversing Rejections, 37 CFR 1.132	717.01(b)(1)	Evaluation of Declarations or Affidavits under 37 CFR 1.130(b)
716.01	Generally Applicable Criteria		
716.01(a)	Objective Evidence of Nonobviousness	717.01(b)(2)	Determining if the Subject Matter of the Intervening Disclosure is the Same as the Subject Matter of the Inventor-Originated Prior Public Disclosure
716.01(b)	Nexus Requirement and Evidence of Nonobviousness		
716.01(c)	Probative Value of Objective Evidence		

717.01(c) Who May Make Affidavit or Declaration; Formal Requirements of Affidavits and Declarations

717.01(d) U.S. Patent or Application Publication Claiming Same Invention

717.01(e) Passed Upon (or Decided by) by Primary Examiner

717.01(f) Seasonable (or Timely) Presentation

717.02 Prior Art Exception for Commonly Owned or Joint Research Agreement Subject Matter under AIA 35 U.S.C. 102(b)(2)(C)

717.02(a) Invoking the Prior Art Exception under 35 U.S.C. 102(b)(2)(C)

717.02(b) Evaluating Whether the Prior Art Exception under AIA 35 U.S.C. 102(b)(2)(C) is Properly Invoked

717.02(c) Examination Procedure With Respect to the Prior Art Exception under AIA 35 U.S.C. 102(b)(2)(C)

717.02(d) Form Paragraphs With Respect to the Prior Art Exception under AIA 35 U.S.C. 102(b)(2)(C)

718 **Affidavit or Declaration to Disqualify Commonly Owned Patent as Prior Art, 37 CFR 1.131(c)**

719 **File Wrapper**

719.01 Papers in Image File Wrapper

719.02 Residence of Inventor Changed

719.03 Classification During Examination

719.04 Index of Claims

719.05 Field of Search

720 **Public Use Proceedings**

721-723 **[Reserved]**

724 **Trade Secret, Proprietary, and Protective Order Materials**

724.01 Completeness of the Patent File Wrapper

724.02 Method of Submitting Trade Secret, Proprietary, and/or Protective Order Materials

724.03 Types of Trade Secret, Proprietary, and/or Protective Order Materials Submitted Under MPEP § 724.02

724.04 Office Treatment and Handling of Materials Submitted Under MPEP § 724.02

724.04(a) Materials Submitted in an Application Covered by 35 U.S.C. 122

724.04(b) Materials Submitted in Reissue Applications Open to the Public Under 37 CFR 1.11(b)

724.04(c) Materials Submitted in Reexamination File Open to the Public Under 37 CFR 1.11(d)

724.05 Petition To Expunge Information or Copy of Papers in Application File

724.06 Handling of Petitions To Expunge Information or Copy of Papers in Application File

701 Statutory Authority for Examination [R-10.2019]

35 U.S.C. 131 Examination of application.

The Director shall cause an examination to be made of the application and the alleged new invention; and if on such examination it appears that the applicant is entitled to a patent under the law, the Director shall issue a patent therefor.

The main conditions precedent to the grant of a patent to an applicant are set forth in [35 U.S.C. 101](#), [102](#), [103](#), and [112](#).

35 U.S.C. 101 Inventions patentable.

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

See MPEP §§ [2104](#) – [2109](#) and [2157](#) for a discussion of inventions that are not considered to be patentable under [35 U.S.C. 101](#).

35 U.S.C. 100 Definitions.

[Editor Note: 35 U.S.C. 100(e)-(j) as set forth below are only applicable to patent applications and patents subject to the first inventor to file provisions of the AIA (35 U.S.C. 100 (note)). See pre-AIA 35 U.S.C. 100(e) for paragraph (e) as applicable to patent applications and patents not subject to the first inventor to file provisions of the AIA.]

When used in this title unless the context otherwise indicates -

(a) The term “invention” means invention or discovery.

(b) The term “process” means process, art, or method, and includes a new use of a known process, machine, manufacture, composition of matter, or material.

invention under [35 U.S.C. 102\(a\)\(2\)](#) if: (1) The subject matter disclosed was obtained directly or indirectly from the inventor or a joint inventor; or (2) the subject matter disclosed had, before such subject matter was effectively filed, been publicly disclosed by the inventor or a joint inventor or another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor.

The Office has provided a mechanism in [37 CFR 1.130](#) for filing an affidavit or declaration to establish that a disclosure made no earlier than one year before the effective filing date of the claimed invention is not prior art under [35 U.S.C. 102\(a\)](#) due to an exception in [35 U.S.C. 102\(b\)](#). See [MPEP § 717.01 et seq.](#) for more information on declarations filed under [37 CFR 1.130](#).

Additionally, [35 U.S.C. 102\(b\)\(2\)\(C\)](#) provides that a disclosure made in a U.S. patent, U.S. patent application publication, or WIPO published application shall not be prior art to a claimed invention under [35 U.S.C. 102\(a\)\(2\)](#) if, not later than the effective filing date of the claimed invention, the subject matter disclosed and the claimed invention were owned by the same person or subject to an obligation of assignment to the same person. This provision replaces the exception in [pre-AIA 35 U.S.C. 103\(c\)](#) that applied only in the context of an obviousness analysis under [35 U.S.C. 103](#) to prior art that was commonly owned at the time the claimed invention was made, and which qualified as prior art only under [pre-AIA 35 U.S.C. 102\(e\), \(f\), and/or \(g\)](#). Thus, the AIA provides that certain prior patents and published patent applications of co-workers and collaborators are not prior art either for purposes of determining novelty ([35 U.S.C. 102](#)) or nonobviousness ([35 U.S.C. 103](#)). See [MPEP § 717.02 et seq.](#) for more information on the prior art exclusion under [35 U.S.C. 102\(b\)\(2\)\(C\)](#).

The AIA does not define the term “disclosure,” and [35 U.S.C. 102\(a\)](#) does not use the term “disclosure.” [35 U.S.C. 102\(b\)\(1\)](#) and [102\(b\)\(2\)](#), however, each state conditions under which a “disclosure” that otherwise falls within [35 U.S.C. 102\(a\)\(1\)](#) or [102\(a\)\(2\)](#) is not prior art under [35 U.S.C. 102\(a\)\(1\)](#) or [102\(a\)\(2\)](#). Thus, the Office is treating the term “disclosure” as a generic expression intended to

encompass the documents and activities enumerated in [35 U.S.C. 102\(a\)](#) (i.e., being patented, described in a printed publication, in public use, on sale, or otherwise available to the public, or being described in a U.S. patent, U.S. patent application publication, or WIPO published application).

For more information on the provisions of [35 U.S.C. 102](#) in general, see [MPEP § 2150 et seq.](#) For more information on the prior art exceptions under [35 U.S.C. 102\(b\)\(1\)](#), see [MPEP § 2153 et seq.](#) For more information on the prior art exceptions under [35 U.S.C. 102\(b\)\(2\)](#), see [MPEP § 2154.02 et seq.](#)

717.01 Affidavit or Declaration Under 37 CFR 1.130 [R-10.2019]

*[Editor Note: This MPEP section is **only applicable** to applications subject to the first inventor to file provisions of the AIA. See [35 U.S.C. 100 \(note\)](#) and [MPEP § 2159](#). For applications subject to [pre-AIA 35 U.S.C. 102](#) and [pre-AIA 35 U.S.C. 103](#), see [MPEP § 715 et seq.](#) for affidavits or declarations under [37 CFR 1.131\(a\)](#), [MPEP § 718](#) for affidavits or declarations under [37 CFR 1.131\(c\)](#), and [MPEP § 716.10](#) for affidavits or declarations of attribution under [37 CFR 1.132](#).]*

37 CFR 1.130 Affidavit or declaration of attribution or prior public disclosure under the Leahy-Smith America Invents Act.

(a) *Affidavit or declaration of attribution.* When any claim of an application or a patent under reexamination is rejected, the applicant or patent owner may submit an appropriate affidavit or declaration to disqualify a disclosure as prior art by establishing that the disclosure was made by the inventor or a joint inventor, or the subject matter disclosed was obtained directly or indirectly from the inventor or a joint inventor.

(b) *Affidavit or declaration of prior public disclosure.* When any claim of an application or a patent under reexamination is rejected, the applicant or patent owner may submit an appropriate affidavit or declaration to disqualify a disclosure as prior art by establishing that the subject matter disclosed had, before such disclosure was made or before such subject matter was effectively filed, been publicly disclosed by the inventor or a joint inventor or another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor. An affidavit or declaration under this paragraph must identify the subject matter publicly disclosed and provide the date such subject matter was publicly disclosed by the inventor or a joint inventor or another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor.

(1) If the subject matter publicly disclosed on that date was in a printed publication, the affidavit or declaration must be accompanied by a copy of the printed publication.

(2) If the subject matter publicly disclosed on that date was not in a printed publication, the affidavit or declaration must describe the subject matter with sufficient detail and particularity to determine what subject matter had been publicly disclosed on that date by the inventor or a joint inventor or another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor.

(c) *When this section is not available*. The provisions of this section are not available if the rejection is based upon a disclosure made more than one year before the effective filing date of the claimed invention. The provisions of this section may not be available if the rejection is based upon a U.S. patent or U.S. patent application publication of a patented or pending application naming another inventor, the patent or pending application claims an invention that is the same or substantially the same as the applicant's or patent owner's claimed invention, and the affidavit or declaration contends that an inventor named in the U.S. patent or U.S. patent application publication derived the claimed invention from the inventor or a joint inventor named in the application or patent, in which case an applicant or a patent owner may file a petition for a derivation proceeding pursuant to § 42.401 *et seq.* of this title

(d) *Applications and patents to which this section is applicable*. The provisions of this section apply to any application for patent, and to any patent issuing thereon, that contains, or contained at any time:

(1) A claim to a claimed invention that has an effective filing date as defined in § 1.109 that is on or after March 16, 2013; or

(2) A specific reference under 35 U.S.C. 120, 121, 365(c), or 386(c) to any patent or application that contains, or contained at any time, a claim to a claimed invention that has an effective filing date as defined in § 1.109 that is on or after March 16, 2013.

37 CFR 1.130 provides a mechanism for filing an affidavit or declaration to establish that a disclosure is not prior art in accordance with 35 U.S.C. 102(b). In other words, 37 CFR 1.130, like 37 CFR 1.131 and 37 CFR 1.132, provides a mechanism for the submission of evidence to disqualify a disclosure as prior art or otherwise traverse a rejection. An applicant's or patent owner's compliance with 37 CFR 1.130 means that the applicant or patent owner is entitled to have the evidence considered in determining the patentability of the claim(s) at issue. It does not mean that the applicant or patent owner is entitled as a matter of right to have the rejection of, or objection to, the claim(s) withdrawn. See Changes To Implement the Patent Business Goals, 65 FR 54604, 54640 (Sept. 8, 2000) (discussing

procedural nature of 37 CFR 1.131 and 37 CFR 1.132).

37 CFR 1.130(a) provides that when any claim of an application or a patent under reexamination is rejected, the applicant or patent owner may submit an appropriate affidavit or declaration to disqualify a disclosure as prior art by establishing that the disclosure was made by the inventor or a joint inventor, or the subject matter disclosed was obtained directly or indirectly from the inventor or a joint inventor. 37 CFR 1.130(a) pertains to the provisions of subparagraph (A) of 35 U.S.C. 102(b)(1) and 102(b)(2).

37 CFR 1.130(b) provides that when any claim of an application or a patent under reexamination is rejected, the applicant or patent owner may submit an appropriate affidavit or declaration to disqualify a disclosure as prior art by establishing that the subject matter disclosed had, before such disclosure was made or before such subject matter was effectively filed, been publicly disclosed by the inventor or a joint inventor or another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor. 37 CFR 1.130(b) pertains to the provisions of subparagraph (B) of 35 U.S.C. 102(b)(1) and 102(b)(2).

37 CFR 1.130(c) provides that the provisions of 37 CFR 1.130 are not available if the rejection is based upon a disclosure made more than one year before the effective filing date of the claimed invention. A disclosure made more than one year before the effective filing date of the claimed invention is prior art under 35 U.S.C. 102(a)(1) and does not meet the requirements for exception under 35 U.S.C. 102(b)(1).

37 CFR 1.130(d) provides that the provisions of 37 CFR 1.130 apply to any application for patent, and to any patent issuing thereon, that is subject to 35 U.S.C. 102 and 103.

I. SITUATIONS WHERE 37 CFR 1.130 AFFIDAVITS OR DECLARATIONS CAN BE USED

Affidavits or declarations under 37 CFR 1.130 may be used, for example:

(A) When a claim is under a prior art rejection, to show a disclosure (e.g., reference, activity, or portion thereof) that is relied upon as prior art under [35 U.S.C. 102\(a\)\(1\)](#) or [\(2\)](#) is eligible to be excepted under [35 U.S.C. 102\(b\)\(1\)](#) or [\(2\)](#) by establishing that the disclosure was made by the inventor or a joint inventor, or the subject matter disclosed was obtained directly or indirectly from the inventor or a joint inventor, subject to certain restrictions, as discussed below in subsection II.

(B) When a claim is under a prior art rejection, to show a disclosure (e.g., reference, activity, or portion thereof) that is relied upon as prior art under [35 U.S.C. 102\(a\)\(1\)](#) or [\(2\)](#) is eligible to be excepted under [35 U.S.C. 102\(b\)\(1\)](#) or [\(2\)](#) by establishing that the subject matter disclosed had, before such disclosure was made or before such subject matter was effectively filed, been publicly disclosed by the inventor or a joint inventor or another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor, subject to certain restrictions, as discussed below in subsection II.

If effective, the showing that the disclosure is not prior art under [35 U.S.C. 102\(a\)](#) will make the disclosure unavailable to be applied in a rejection under either [35 U.S.C. 102\(a\)](#) or [35 U.S.C. 103](#). For example, an applicant or patent owner may overcome a [35 U.S.C. 103](#) rejection based on a combination of references by showing that the subject matter disclosed in only one of the references (or portion thereof) applied in the rejection is not prior art. The U.S. patents or U.S. patent application publications may continue to be applied under the judicially created doctrine of nonstatutory double patenting, statutory double patenting and may serve as evidence of the level of ordinary skill or evidence relevant to an enablement inquiry.

II. SITUATIONS WHERE 37 CFR 1.130 AFFIDAVITS OR DECLARATIONS ARE INAPPROPRIATE

An affidavit or declaration under [37 CFR 1.130](#) is not appropriate in the following situations:

(A) When the disclosure that is applied in a rejection is:

(1) a U.S. patent or U.S. patent application publication that

(a) claims an invention that is the same or substantially the same as the applicant's or patent owner's claimed invention, and

(b) the affidavit or declaration contends that an inventor named in the U.S. patent or U.S. patent application publication derived the claimed invention from the inventor or a joint inventor named in the application or patent under examination. In this case, an applicant or patent owner may file a petition for a derivation proceeding pursuant to [37 CFR 42.401](#) *et seq.*

The provisions of [37 CFR 1.130](#), however, would be available if:

(1) The rejection is based upon a disclosure other than a U.S. patent or U.S. patent application publication (such as non-patent literature or a foreign patent document);

(2) the rejection is based upon a U.S. patent or U.S. patent application and the patent or pending application did not claim an invention that is the same or substantially the same as the applicant's claimed invention; or

(3) the rejection is based upon a U.S. patent or U.S. patent application and while the patent or pending application does claim an invention that is the same or substantially the same as the applicant's claimed invention, the affidavit or declaration under [37 CFR 1.130](#) does not contend that an inventor named in the U.S. patent or U.S. patent application publication derived the claimed invention from the inventor or a joint inventor named in the application or patent under examination (e.g., an affidavit or declaration under [37 CFR 1.130](#) would be available if instead of alleging derivation, the affidavit or declaration under [37 CFR 1.130](#) contends that the subject matter disclosed had, before such disclosure was made or before such subject matter was effectively filed, been publicly disclosed by the inventor or a joint inventor or by another who obtained the subject matter disclosed directly or indirectly from the inventor or a joint inventor).

(B) If the rejection is based upon a **disclosure made more than one year before the effective filing date of the claimed invention**. A disclosure made more than one year before the effective filing date of the claimed invention is prior art under [35 U.S.C. 102\(a\)\(1\)](#) and may not be excepted under [35 U.S.C. 102\(b\)\(1\)](#). Note that the provisions of [37 CFR](#)

[1.130](#) are available to overcome a rejection under [35 U.S.C. 102\(a\)\(2\)](#) which is based on subject matter in an application or patent that was effectively filed, but not published or made publicly available, more than one year before the effective filing date of the claimed invention under examination, where the subject matter disclosed was obtained directly or indirectly from the inventor or a joint inventor. See also, [MPEP § 2155.06](#).

III. SITUATIONS WHERE 37 CFR 1.130(a) AFFIDAVITS OR DECLARATIONS ARE NOT REQUIRED

(A) A declaration under [37 CFR 1.130\(a\)](#) is not required when a public disclosure, subject to the exceptions of [35 U.S.C. 102\(b\)\(1\)\(A\)](#), is by one or more joint inventor(s) or the entire inventive entity of the application under examination and does not name anyone else. For example, if an application names A, B, and C as the inventive entity, a journal publication names as authors A and B, and the publication date is one year or less before the effective filing date of the claimed invention, then the publication should not be applied in a prior art rejection because it is apparent that the disclosure is a grace period disclosure. Where a disclosure involves a patent document, a declaration under [37 CFR 1.130\(a\)](#) is not required where the inventive entity of the patent document, subject to the exceptions of [35 U.S.C. 102\(b\)\(1\)\(A\)](#), only includes one or more joint inventor(s) or the entire inventive entity of the application under examination. For example, if the application names A, B, and C as the inventive entity, the patent document names A and B as the inventive entity, and the public availability date of the patent document is one year or less before the effective filing date of the claimed invention under examination, then the patent document should not be applied in a prior art rejection under [35 U.S.C. 102\(a\)\(1\)](#) because it is apparent that the patent document disclosure is a grace period disclosure.

(B) A declaration under [37 CFR 1.130\(a\)](#) is not required when the inventive entity of a U.S. patent, U.S. patent application publication, or a WIPO publication that designates the United States, subject to the exceptions of [35 U.S.C. 102\(b\)\(2\)\(A\)](#), only includes one or more joint inventor(s), but not the entire inventive entity, of the application under examination. For example, if the application under examination names as the inventive entity A, B, and

C, and the [35 U.S.C. 102\(a\)\(2\)](#) reference names A and B as the inventive entity, then the reference should not be applied in a prior art rejection because it is apparent that the subject matter disclosed was obtained from one or more members of the inventive entity, either directly or indirectly. The exceptions under [35 U.S.C. 102\(b\)\(1\)\(A\)](#) and [35 U.S.C. 102\(b\)\(2\)\(A\)](#) both have to do with disclosures of material that originated with the inventor or a joint inventor. However, the [35 U.S.C. 102\(b\)\(2\)\(A\)](#) exception is not limited to the grace period.

(C) A declaration under [37 CFR 1.130\(a\)](#) is not required if the specification of the application under examination identifies the disclosure or the subject matter disclosed that is subject to the exceptions of [35 U.S.C. 102\(b\)\(1\)\(A\)](#) or [35 U.S.C. 102\(b\)\(2\)\(A\)](#) as having been made by or having originated from one or more members of the inventive entity of the application under examination. See [37 CFR 1.77\(b\)\(6\)](#) and [MPEP § 2153.01\(a\)](#).

IV. PRIOR ART DATE OF THE DISCLOSURE UNDER AIA 35 U.S.C. 102(a)

For determining the prior art date of a disclosure under AIA [35 U.S.C. 102\(a\)\(1\)](#), see [MPEP § 2152.02\(a\)](#) *et seq.*

For determining the prior art date of a disclosure under AIA [35 U.S.C. 102\(a\)\(2\)](#), see [MPEP § 2154.01](#) *et seq.*

V. FORM PARAGRAPHS

¶ 7.67.aia Affidavit or Declaration Under 37 CFR 1.130: Effective to Disqualify a Reference as Prior Art Via 35 U.S.C. 102(b)

The [1] under [37 CFR 1.130](#) [2] filed on [3] is sufficient to overcome the rejection of claim [4] based on [5]. [6]

Examiner Note:

1. This form paragraph should only be used in an application filed on or after March 16, 2013, where the claims are being examined under [35 U.S.C. 102/103](#) as amended by the Leahy-Smith America Invents Act. This form paragraph must be preceded by form paragraph 7.03.aia.
2. In bracket 1, insert either --affidavit-- or --declaration--.
3. In bracket 2, insert either --(a)-- or --(b)--.
4. In bracket 3, insert the filing date of the affidavit or declaration
5. In bracket 4, insert the affected claim or claims.

EXHIBIT 14

Practitioner's Docket No. 030258-079361-P

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Anthony John Iafrate et al Confirmation No.: TBD
Serial No.: TBD Art Unit: TBD
Filed: Herewith Examiner: TBD
For: METHODS OF PREPARING NUCLEIC ACIDS FOR SEQUENCING

VIA EFS-WEB

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

GENERAL TRANSMITTAL & CERTIFICATE OF EFS SUBMISSION

Pursuant to 37 C.F.R. § 1.8(a)(i)(1)(C) I hereby certify that the following documents:

1. General Transmittal & Certificate of EFS Submission (1 pg.);
2. PTO/SB/16 – Provisional Application Cover Sheet (4 pp.);
3. Provisional Application (69 pp.); Specification (54 pp.); Claims (9 pp.); Abstract (1 pg.); Figs. 1A-3 (5 pp.);
4. EFS Fee Transmittal (2 pp.)

are being submitted to the Patent and Trademark Office via the Office Electronic Filing System in accordance with § 1.6(a)(4) on **January 27, 2014**.

AUTHORIZATION TO CHARGE/CREDIT DEPOSIT ACCOUNT

The Commissioner is hereby authorized to charge fee deficiencies or credit overpayments associated with the instant filing in above-referenced matter to NIXON PEABODY LLP Deposit Account No. 50-0850.

Date: January 27, 2014

Respectfully submitted,

Customer No.: 50828

/David S. Resnick/
David S. Resnick (Reg. No. 34,235)
NIXON PEABODY LLP
100 Summer Street
Boston, MA 02110-2131
Tel. (617) 345-6057
Fax (617) 345-1300
E-mail: bostonpatent@nixonpeabody.com

ADX - 000365412

Electronic Patent Application Fee Transmittal				
Application Number:				
Filing Date:				
Title of Invention:		Methods Of Preparing Nucleic Acids For Sequencing		
First Named Inventor/Applicant Name:		Anthony John lafrate		
Filer:		David S. Resnick		
Attorney Docket Number:		030258-079361-P		
Filed as Small Entity				
Provisional Filing Fees				
Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Basic Filing:				
Provisional Application Filing Fee	2005	1	130	130
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

ADX - 000365413

Description	Fee Code	Quantity	Amount	Sub-Total in USD(\$)
Miscellaneous:				
Total in USD (\$)				130

Electronic Acknowledgement Receipt

EFS ID:	18033922
Application Number:	61931959
International Application Number:	
Confirmation Number:	3887
Title of Invention:	Methods Of Preparing Nucleic Acids For Sequencing
First Named Inventor/Applicant Name:	Anthony John lafrate
Customer Number:	50828
Filer:	David S. Resnick
Filer Authorized By:	
Attorney Docket Number:	030258-079361-P
Receipt Date:	27-JAN-2014
Filing Date:	
Time Stamp:	16:25:08
Application Type:	Provisional

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	\$ 130
RAM confirmation Number	3576
Deposit Account	500850
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees) Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)	

ADX - 000365415

Charge any Additional Fees required under 37 C.F.R. Section 1.19 (Document supply fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Provisional Cover Sheet (SB16)	20140127_ProvisionalCoversheet_030258_079361_P.pdf	1477527	no	4
			e94336d9ec435fe8b93c8f26ed9a26bab6622abc		
Warnings:					
Information:					
2		20140127_ProvisionalApplication_030258_079361_P.pdf	311976	yes	64
			fa090631574fc83006f54e62bdb73c757738e52b		
	Multipart Description/PDF files in .zip description				
	Document Description		Start	End	
	Specification		1	54	
	Claims		55	63	
	Abstract		64	64	
Warnings:					
Information:					
3	Drawings-only black and white line drawings	20140127_ProvisionalApplicationFigures_030258_079361_P.pdf	85537	no	5
			938feacc9d9d839d0a40d60d90d73e3fea82bd3b		
Warnings:					
Information:					
4	Miscellaneous Incoming Letter	20140127_EFSTransmittalProvisionalApplication_030258_079361_P.pdf	90879	no	1
			17d03677b9aa5681e35ce2032240acfad0cc0832		
Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	29633	no	2
			04b84a735e2927f7250933fe3ac612ed645108e9		
Warnings:					
Information:					
Total Files Size (in bytes):			1995552		

ADX - 000365416

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.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

Provisional Application for Patent Cover Sheet					
This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c)					
Inventor(s)					
Inventor 1					Remove
Given Name	Middle Name	Family Name	City	State	Country i
Anthony	John	Iafrate	Boston	MA	US
Inventor 2					Remove
Given Name	Middle Name	Family Name	City	State	Country i
Long	Phi	Le	Boston	MA	US
Inventor 3					Remove
Given Name	Middle Name	Family Name	City	State	Country i
Zongli		Zheng	Boston	MA	US
Inventor 4					Remove
Given Name	Middle Name	Family Name	City	State	Country i
Jason		Myers	Golden	CO	US
Inventor 5					Remove
Given Name	Middle Name	Family Name	City	State	Country i
Joshua		Stahl	Boulder	CO	US
All Inventors Must Be Listed – Additional Inventor Information blocks may be generated within this form by selecting the Add button.					Add
Title of Invention		METHODS OF PREPARING NUCLEIC ACIDS FOR SEQUENCING			
Attorney Docket Number (if applicable)		030258-079361-P			
Correspondence Address					
Direct all correspondence to (select one):					
☒ The address corresponding to Customer Number			☐ Firm or Individual Name		
Customer Number			50828		

Doc Code: TR.PRV

Document Description: Provisional Cover Sheet (SB16)

PTO/SB/16 (11-08)

Approved for use through 05/31/2015. 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

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the invention was made by an agency of the United States Government. The U.S. Government agency name is:

☐ Yes, the invention was under a contract with an agency of the United States Government. The name of the U.S. Government agency and Government contract number are:

Doc Code: TR-PROV

Document Description: Provisional Cover Sheet (SB16)

PTO/SB/16 (11-08)

Approved for use through 05/31/2015. 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

Entity Status**Applicant asserts small entity status under 37 CFR 1.27 or applicant certifies micro entity status under 37 CFR 1.29**

- ☒ Applicant asserts small entity status under 37 CFR 1.27
- ☐ Applicant certifies micro entity status under 37 CFR 1.29. Applicant must attach form PTO/SB/15A or B or equivalent.
- ☐ No

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 USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

Signature

Please see 37 CFR 1.4(d) for the form of the signature.

Signature	/David S. Resnick/			Date (YYYY-MM-DD)	2014-01-27
First Name	David S.	Last Name	Resnick	Registration Number (If appropriate)	34235

This collection of information is required by 37 CFR 1.51. 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 8 hours 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. **This form can only be used when in conjunction with EFS-Web. If this form is mailed to the USPTO, it may cause delays in handling the provisional application.**

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.

METHODS OF PREPARING NUCLEIC ACIDS FOR SEQUENCING

TECHNICAL FIELD

[0001] The technology described herein relates to methods of preparing and analyzing nucleic acids.

BACKGROUND

[0002] Target enrichment prior to next-generation sequencing is more cost-effective than whole genome, whole exome, and whole transcriptome sequencing and therefore more practical for broad implementation; both for research discovery and clinical applications. For example, high coverage depth afforded by target enrichment approaches enables a wider dynamic range for allele counting (in gene expression and copy number assessment) and detection of low frequency mutations, a critical feature for evaluating somatic mutations in cancer. Examples of current enrichment protocols for next generation sequencing include hybridization-based capture assays (TruSeq Capture, Illumina; SureSelect Hybrid Capture, Agilent) and polymerase chain reaction (PCR)-based assays (HaloPlex, Agilent; AmpliSeq, Ion Torrent; TruSeq Amplicon, Illumina; emulsion/digital PCR, Raindance). Hybridization-based approaches capture not only the targeted sequences covered by the capture probes but also near off-target bases that consume sequencing capacity. In addition, these methods are relatively time-consuming, labor-intensive, and suffer from a relatively low level of specificity.

SUMMARY

[0003] Aspects of the technology disclosed herein relate to methods for preparing and analyzing nucleic acids. In some embodiments, methods for preparing nucleic acids for sequence analysis (*e.g.*, using next-generating sequencing) are provided herein. In some embodiments, technology described herein is directed to methods of determining nucleotide sequences of nucleic acids. In some embodiments, the methods described herein relate to enriching target nucleic acids prior to sequencing.

[0004] Aspects of the technology disclosed herein relate to methods of determining the nucleotide sequence contiguous to a known target nucleotide sequence. In some embodiments, the methods involve (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with an initial target-specific primer under hybridization conditions; (b) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid molecule as a template; (c) contacting the product of step (b) with a population of tailed

random primers under hybridization conditions; (d) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template; (e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer; (f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer; and (g) sequencing the amplified portion from step (f) using a first and second sequencing primer. In some embodiments, the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides. In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (e), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence identical to the tailed random primer. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

[0005] In some embodiments, the methods involve (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions; (b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template; (c) contacting the product of step (b) with an initial target-specific primer under hybridization conditions; (d) performing a template-dependent extension reaction that is primed by a hybridized initial target-specific primer and that uses the target nucleic acid molecule as a template; (e) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer; (f) amplifying a portion of the amplicon resulting from step (e) with a second tail primer and a second target-specific primer; and (g) sequencing the amplified portion from step (f) using a

first and second sequencing primer. In some embodiments, the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides. In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence identical to the tailed random primer. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

[0006] In some embodiments, the methods further involve a step of contacting the sample and/or products with RNase after extension of the initial target-specific primer. In some embodiments, the tailed random primer can form a hair-pin loop structure. In some embodiments, the initial target-specific primer and the first target-specific primer are identical. In some embodiments, the tailed random primer further comprises a barcode portion comprising 6-12 random nucleotides between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising 6-12 random nucleotides.

[0007] In some embodiments, the methods involve (a) contacting a target nucleic acid molecule comprising the known target nucleotide sequence with a population of tailed random primers under hybridization conditions; (b) performing a template-dependent extension reaction that is primed by a hybridized tailed random primer and that uses the portion of the target nucleic acid molecule downstream of the site of hybridization as a template; (c) amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer; (d) amplifying a portion of the amplicon resulting from step (c) with a second tail primer and a second target-specific primer; and (e) sequencing the amplified portion from step (d) using a first and second sequencing primer. In some embodiments, the population of tailed random primers

comprises single-stranded oligonucleotide molecules having a 5' nucleic acid sequence identical to a first sequencing primer; a middle barcode portion comprising; and a 3' nucleic acid sequence comprising from about 6 to about 12 random nucleotides. In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known target nucleotide sequence of the target nucleic acid at the annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the amplicon resulting from step (c), and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence identical to the tailed random primer. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer. In some embodiments, the each tailed random primer further comprises a spacer nucleic acid sequence between the 5' nucleic acid sequence identical to a first sequencing primer and the 3' nucleic acid sequence comprising about 6 to about 12 random nucleotides. In certain embodiments, the unhybridized primers are removed from the reaction after an extension step. In some embodiments, the second tail primer is nested with respect to the first tail primer by at least 3 nucleotides. In certain embodiments, the first target-specific primer further comprises a 5' tag sequence portion comprising a nucleic acid sequence of high GC content which is not substantially complementary to or substantially identical to any other portion of any of the primers. In some embodiments, the second tail primer is identical to the full-length first sequencing primer. In certain embodiments, the portions of the target-specific primers that specifically anneal to the known target will anneal specifically at a temperature of about 65°C in a PCR buffer. In some embodiments, the sample comprises genomic DNA. In some embodiments, the sample comprises RNA and the method further comprises a first step of subjecting the sample to a reverse transcriptase regimen. In certain embodiments, the nucleic acids present in the sample have not been subjected to shearing or digestion. In some embodiments, the sample comprises single-stranded gDNA or cDNA. In certain embodiments, the reverse transcriptase regimen comprises the use of random hexamers. In some embodiments, a gene rearrangement comprises the known target sequence. In certain embodiments, the gene rearrangement is present in a nucleic acid selected from the group consisting of: genomic DNA; RNA; and

cDNA. In some embodiments, the gene rearrangement comprises an oncogene. In certain embodiments, the gene rearrangement comprises a fusion oncogene. In some embodiments, the nucleic acid product is sequenced by a next-generation sequencing method. In certain embodiments, the next-generation sequencing method comprises a method selected from the group consisting of: Ion Torrent, Illumina, SOLiD, 454; Massively Parallel Signature Sequencing solid-phase, reversible dye-terminator sequencing; and DNA nanoball sequencing. In certain embodiments, the first and second sequencing primers are compatible with the selected next-generation sequencing method. In some embodiments, the method comprises contacting the sample, or separate portions of the sample, with a plurality of sets of first and second target-specific primers. In certain embodiments, the method comprises contacting a single reaction mixture comprising the sample with a plurality of sets of first and second target-specific primers. In some embodiments, the plurality of sets of first and second target-specific primers specifically anneal to known target nucleotide sequences comprised by separate genes. In certain embodiments, at least two sets of first and second target-specific primers specifically anneal to different portions of a known target nucleotide sequence. In some embodiments, at least two sets of first and second target-specific primers specifically anneal to different portions of a single gene comprising a known target nucleotide sequence. In certain embodiments, at least two sets of first and second target-specific primers specifically anneal to different exons of a gene comprising a known nucleotide target sequence. In some embodiments, the plurality of first target-specific primers comprise identical 5' tag sequence portions. In certain embodiments, each tailed random primer in a population of tailed random primers further comprises an identical sample barcoding portion. In some embodiments, multiple samples are each contacted with a separate population of tailed random primers with a sample barcoding portion; wherein each population of tailed random primers has a distinct sample barcoding portion; and wherein the samples are pooled after step (b). In certain embodiments, each amplification step comprises a set of cycles of a PCR amplification regimen from 5 cycles to 20 cycles in length. In some embodiments, the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of from about 61 to 72 °C. In some embodiments, the target-specific primers and the tail primers are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of about 65 °C. In certain embodiments, the target nucleic acid molecule is from a sample, optionally which is a biological sample obtained from a subject.

In some embodiments, the sample is obtained from a subject in need of treatment for a disease associated with a genetic alteration. In certain embodiments, the disease is cancer. In some embodiments, the sample comprises a population of tumor cells. In certain embodiments, the sample is a tumor biopsy. In some embodiments, the cancer is lung cancer. In certain embodiments, a disease-associated gene comprises the known target sequence. In some embodiments, a gene rearrangement product in the sample comprises the known target sequence. In certain embodiments, the gene rearrangement product is an oncogene.

[0008] Aspects of the technology disclosed herein relate to methods of preparing nucleic acids for analysis. In some embodiments, the methods involve method (a) contacting a nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer; and (b) contacting a nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers, in which an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers. In some embodiments, the target nucleic acid is a ribonucleic acid. In certain embodiments, the target nucleic acid is a deoxyribonucleic acid. In some embodiments, steps (a) and (b) are performed sequentially. In certain embodiments, the nucleic acid template in step (a) comprises an extension product resulting from the hybridization and extension of the at least one of the plurality of different primers in step (b). In some embodiments, the nucleic acid template in step (b) comprises an extension product resulting from the hybridization and extension of the target-specific primer in step (a). In certain embodiments, the target nucleic acid is a messenger RNA encoded from a chromosomal segment that comprises a genetic rearrangement. In some embodiments, the target nucleic acid is a chromosomal segment that comprises a portion of a genetic rearrangement. In certain embodiments, the genetic rearrangement is an inversion, deletion, or translocation. In some embodiments, the methods further involve amplifying the extension product. In certain embodiments, the methods further involve contacting the extension product or amplified extension product with an immobilized oligonucleotide under conditions in which hybridization occurs between the extension product and immobilized

oligonucleotide. In certain embodiments, the target nucleic acid comprises a target portion having a known sequence and a flanking portion having an unknown sequence. In some embodiments, different hybridization sequences are complementary to the flanking portion. In certain embodiments, the target-specific hybridization sequence is complementary to the target portion. In some embodiments, the target-specific primer further comprises, 5' to the target-specific hybridization sequence, at least one of an index sequence, a barcode sequence and an adaptor sequence. In certain embodiments, the common sequence comprises at least one of an index sequence, barcode sequence and an adaptor sequence. In some embodiments, the adaptor sequence is a cleavable adaptor sequence for immobilizing oligonucleotides in a flow cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIGS. 1A and B depict non-limiting embodiments of a work flow for amplifying and sequencing target nucleic acids that are flanked by a 3' unknown fusion partner, as described herein.

[0010] FIGS. 2 A and B depict non-limiting embodiments of a work flow for amplifying sequencing target nucleic acids that are flanked by a 5' unknown fusion partner, as described herein.

[0011] FIG. 3 depicts a non-limiting embodiment of a work flow for amplifying sequencing target nucleic acids that are flanked by a 3' unknown fusion partner using an oligonucleotide comprising a hairpin, as described herein.

DETAILED DESCRIPTION

[0012] Aspects of the technology disclosed herein relate to methods for preparing and analyzing nucleic acids. In some embodiments, methods provided herein are useful for determining unknown nucleotide sequences contiguous to (adjacent to) a known target nucleotide sequence. Traditional sequencing methods generate sequence information randomly (*e.g.* “shotgun” sequencing) or between two known sequences which are used to design primers. In contrast, methods described herein, in some embodiments, allow for determining the nucleotide sequence (*e.g.* sequencing) upstream or downstream of a single region of known sequence with a high level of specificity and sensitivity. Accordingly, in some embodiments, methods provided herein are useful for determining the sequence of fusions (*e.g.*, fusion mRNAs) that result from gene arrangements (*e.g.*, rearrangements that give rise to cancer or other disorders). In some

embodiments, methods provided herein for preparing nucleic acids for analysis (*e.g.*, for sequencing) involve a first round of extension using a target-specific primer that targets a known sequence of a target nucleic acid (*e.g.*, a known sequence of a 1st gene) followed by a second round of extension that involves the use of a heterogeneous population of tailed primers (*e.g.*, tailed random primers) that include tailed primers that have hybridization sequences that are complementary with an unknown sequence adjacent to the known sequence in the target nucleic acid. In some embodiments, the tail region of tailed primers includes barcode or index sequences that facilitate multiplex amplification and enrichment of target nucleic acids.

[0013] In some aspects of the technology disclosed herein methods are provided of preparing nucleic acids for analysis that involve (a) contacting a nucleic acid template comprising a first strand of a target nucleic acid with a complementary target-specific primer that comprises a target-specific hybridization sequence, under conditions to promote template-specific hybridization and extension of the target-specific primer and (b) contacting a nucleic acid template comprising a second strand that is complementary to the first strand of the target nucleic acid with a plurality of different primers that share a common sequence that is 5' to different hybridization sequences, under conditions to promote template-specific hybridization and extension of at least one of the plurality of different primers, in which an extension product is generated to contain both a sequence that is characteristic of the target-specific primer and a sequence that is characteristic of the at least one of the plurality of different primers. In some embodiments, steps (a) and (b) above are performed sequentially. In some embodiments, the nucleic acid template in step (a) comprises an extension product resulting from the hybridization and extension of the at least one of the plurality of different primers in step (b). In some embodiments, the nucleic acid template in step (b) comprises an extension product resulting from the hybridization and extension of the target-specific primer in step (a).

[0014] In some embodiments, methods are provided for preparing nucleic acids that have a target region 5' to an adjacent region (*e.g.*, an adjacent region of unknown sequence). For example, FIGS. 1A-B present schematics of exemplary methods of amplifying target nucleic acids that have a known target region 5' to an adjacent region (*e.g.*, for purposes of sequencing the adjacent region). At step **101**, initial RNA is obtained or provided in a sample and is used as a template. RNA template is exposed to a plurality of tailed primers (*e.g.*, tailed random primers) that comprise a common sequence that is 5' to different hybridization sequences and shared between all of the tailed primers of the population. In some embodiments, at least one primer hybridizes to an RNA molecule and primes a reverse transcriptase reaction to produce a complementary DNA strand. In step **102**, unhybridized oligonucleotides are degraded (*e.g.*,

enzymatically, *e.g.*, by an exonuclease). In step **102**, RNA template is degraded from the complementary DNA strand.

[0015] It should be appreciated that in methods disclosed herein an RNA template may be degraded by any appropriate method, including, for example, by enzymatic degradation (*e.g.*, using RNaseH, Uracyl glycosylase, etc.), by hydrolysis (*e.g.*, by exposing the RNA to relatively high pH conditions (*e.g.*, pH 10, pH 11, pH 12), *etc.* In some embodiments, hydrolyzing RNA by exposure to relatively high pH is advantageous because it is relatively inexpensive (compared with certain other methods, *e.g.*, certain enzymatic methods) and because it destroys both RNA and DNA:RNA hybrids. In some embodiments, RNA is degraded by hydrolysis caused by exposure to relatively high pH conditions at a relatively high temperature (*e.g.*, temperature greater than 60 °C (*e.g.*, 60 °C to 95 °C)) In some embodiments, the use of relatively high temperatures is advantageous because it heat inactivates enzymes used in prior preparative steps (*e.g.*, RT enzymes).

[0016] In step **103**, DNA molecules produced by reverse transcription are contacted by one or more initial target-specific primers which may or may not be the same as the first target-specific primer. In step **104**, hybridization of the initial target-specific primer to a portion of the target nucleic acid primes an extension reaction using a DNA molecule as a template to produce a complementary DNA strand. Extension products are purified in step **105**.

[0017] In step **106**, DNA molecules are contacted by a first target-specific primer and a first tail primer. The first target-specific primer hybridizes to a portion of the target nucleic acid. In some embodiments, pools of different first target-specific primers can be used that hybridize to different portions of a target nucleic acid. In some embodiments, use of different target specific primers can be advantageous because it allows for generation of different extension products having overlapping but staggered sequences relative to a target nucleic acid. In some embodiments, different extension products can be sequenced to produce overlapping sequence reads. In some embodiments, overlapping sequence reads can be evaluated to assess accuracy of sequence information, fidelity of nucleic acid amplification, and/or to increase confidence in detecting mutations, such as detecting locations of chromosomal rearrangements (*e.g.*, fusion breakpoints). In some embodiments, pools of different first target-specific primers can be used that hybridize to different portions of different target nucleic acids present in sample. In some embodiments, use of pools of different target-specific primers is advantageous because it facilitates processing (*e.g.*, amplification) and analysis of different target nucleic acids in parallel. In some embodiments, up to 2, up to 3, up to 4, up to 5, up to 6, up to 7, up to 8, up to 9, up to 10, up to 15, up to 20, up to 100 or more pools of different first target-specific primers are used. In

some embodiments, 2 to 5, 2 to 10, 5 to 10, 5 to 15, 10 to 15, 10 to 20, 10 to 100, 50 to 100, or more pools of different first target-specific primers are used.

[0018] In FIGS. 1A and 1B, a first tail primer hybridizes to at least a portion of a DNA molecule provided by the tail portion of the tailed primer of step **101**. In some embodiments, the first tail primer hybridizes to the common sequence provided by the tail of the one or more primers of step **101**. In some embodiments, a nested target specific primer (nested with respect to the target specific primer of step **103**) is used in step **106**. In some embodiments, a first tail primer may comprise an additional sequence 5' to the hybridization sequence that may include barcode, index, adapter sequences, or sequencing primer sites, for example. In step **107**, hybridization of the first target-specific primer and the first tail nucleic acid molecule in a polymerase chain reaction (PCR). In some embodiments, amplified products are purified in step **108**.

[0019] In some embodiments, an exonuclease (*e.g.*, ExoI) may be used to degrade single-stranded DNA. In some embodiments, an exonuclease is used to degrade ssDNA and the amplified product is processed directly according to steps **109_{A-B}** and **110_{A-B}**, without purification at step **108**.

[0020] In FIG. 1A, at step **109_A**, amplified DNA products of step **107** (*e.g.*, as purified in step **108**) are contacted with a second target-specific primer and a second tail primer. In some embodiments, the second target-specific primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first target-specific primer such that the reactions are nested. In step **110_A**, the amplified DNA products of step **107** (*e.g.*, as purified in step **108**) are amplified by PCR in which the extensions are primed by the second target-specific primer and a second tail primer. In some embodiments, a portion of the amplified product from step **107** is further amplified. In some embodiments, a third primer is used that hybridizes to the common tail in the second target specific primer and adds additional sequences such as a barcodes, adapters, *etc.*

[0021] In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such that the reactions are nested. In such embodiments, a portion of the product from step **106** is amplified. In some embodiments, the second tail primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. Hybridization of the second target-specific primer and the second tail primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction. The

products are purified in reaction **111** and ready for analysis. For example, productions purified in step **111** can be sequenced (*e.g.*, using a next generation sequencing platform.)

[0022] In some embodiments, as depicted in FIG 1B., in step **109_B**, DNA products of step **107** (*e.g.*, as purified in step **108**) are contacted with a second target-specific primer and a second tail primer. The second target-specific primer is further contacted by an additional primer (*e.g.*, a primer having 3' sequencing adapter/index sequences) that hybridizes with the common sequence of the second target-specific primer. In some embodiments the additional primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. In some embodiments, the additional primer is a generic sequencing adapter/index primer. In step **110_B**, the DNA products of step **107** (*e.g.*, as purified in step **108**) are amplified by PCR in which the extensions are primed by the second target-specific primer and a second tail primer. Hybridization of the second target-specific primer, the additional primer, and the second tail primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction. In such embodiments, a portion of the amplified product from step **108** is amplified. The products are purified in reaction **111** and ready for analysis. For example, productions purified in step **111** can be sequenced (*e.g.*, using a next generation sequencing platform).

[0023] In some embodiments, steps **101-104**, **106-107**, and **109-110** are performed consecutively in a single reaction tube without any intervening purification steps. In some embodiments, all of the components involved in steps **101-104**, **106-107**, **109-110** are present at the outset and throughout the reaction. In some embodiments, steps **101-104** are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps **101-104** are present at the outset and throughout the reaction. In some embodiments, steps **106-107** are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps **106-107** are present at the outset and throughout the reaction. In some embodiments, steps **109_A-110_A** or **109_B-110_B** are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps **109_A-110_A** or **109_B-110_B** are present at the outset and throughout the reaction.

[0024] In some embodiments, methods are provided for preparing nucleic acids that have a target region 3' to an adjacent region (*e.g.*, an adjacent region of unknown sequence content). For example, FIG. 2 presents a schematic of an exemplary method of amplifying and sequencing target nucleic acids that have a known target region 3' to an adjacent region. In FIGS. 2A-B, an initial RNA (*e.g.*, a fusion mRNA) is obtained or provided in a sample and is used as a template for the proceeding method. At step **201**, the RNA template is exposed to one or more initial target-specific primers that hybridize to one or more target nucleotide sequences and function to

prime a reverse transcription reaction such that a complementary DNA molecule is produced using the initial RNA as a template. In step **202** the unhybridized primers are degraded (*e.g.*, enzymatically, *e.g.*, by an exonuclease). In step **202**, RNA template is degraded from the complementary DNA strand (*e.g.*, enzymatically, *e.g.*, by RNaseH).

[0025] In step **203**, DNA molecule that were generated by reverse transcription are contacted by a heterogenous population of tailed primers (*e.g.*, tailed random primers). In some embodiments, the tail portion of each of the tailed primers is a shared or common sequence, identical between all primers of the population of tailed primers. In some embodiments, at least one primer comprises a hybridization sequence that is complementary to and hybridizes to the target acid template. In step **204**, tailed primers that are hybridized with the template nucleic acid are extended in a template-dependent extension reaction to produce complementary DNA strands that incorporate the tailed primer sequence and the template sequence. The resulting double stranded DNA product is purified in step **205**.

[0026] In step **206**, DNA products purified in step **205** are contacted with a first target-specific primer and a first tail primer. The first target-specific primer hybridizes to a target sequence (region) of the DNA. The first tail primer hybridizes to a portion of the DNA molecule characteristic of the tail of the tailed primers of step **203**. In some embodiments, the first tail primer hybridizes to the common sequence provided by the tail of the one or more primers of step **203**. In step **207** hybridization of the first target-specific primer and the first tail primer facilitates exponential amplification of a portion of the target nucleic acid molecule in an amplification reaction (*e.g.*, a PCR reaction). In some embodiments, the first tail primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. Amplified products are purified in step **208**.

[0027] In some embodiments, as depicted in FIG. 2A at step **209_A**, DNA molecules produced in step **207** (*e.g.*, as purified in step **208**) are contacted with a second target-specific primer and a second tail primer. In some embodiments, at least a portion of the product from step **209_A** is amplified.

[0028] In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such that the reactions are nested. In some embodiments, at least a portion of the product from step **209_A** is amplified.

[0029] In some embodiments, as depicted in FIG. 2B at step **209_B**, DNA molecules produced in step **207** (*e.g.*, as purified in step **208**) are contacted with a second target-specific primer and a second tail primer wherein the second target-specific primer hybridizes to a target sequence that

is present within the template DNA molecule 3' of the sequence of the first target-specific primer such that the reactions are nested. In some embodiments, at least a portion of the product from step **209_B** is amplified.

[0030] In some embodiments, the second target-specific primer is further contacted by an additional primer (*e.g.*, a primer having 3' sequencing adapter/index sequences) that hybridizes with the common sequence of the second target-specific primer. In some embodiments the additional primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. In some embodiments, the additional primer is a generic sequencing adapter/index primer. In such embodiments, hybridization of the second target-specific primer, the additional primer, and a second tail primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction.

[0031] In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such that the reactions are nested. In some embodiments, at least a portion of the product from step **209_B** is amplified.

[0032] In some embodiments, the second tail primer may comprise additional sequence 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. In step **210**, hybridization of the second target-specific primer and the second tail primer facilitates exponential amplification of a portion of the target nucleic acid molecule in an amplification reaction (*e.g.*, a PCR reaction). The amplification product of step **210** is purified in reaction **211** and ready for analysis. For example, productions purified in step **211** can be sequenced (*e.g.*, using a next generation sequencing platform.)

[0033] In some embodiments, steps **201-204** are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps **201-204** are present at the outset and throughout the reaction. In some embodiments, steps **206-207** are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps **206-207** are present at the outset and throughout the reaction. In some embodiments, steps **209_A-210** or **209_B-210** are performed consecutively in a single reaction tube. In some embodiments, all of the components involve in steps **209_A-210** or **209_B-210** are present at the outset and throughout the reaction.

[0034] In some embodiments, methods provided herein involve use of random sequences as molecular barcodes. In some embodiments, molecular barcodes are built into primers (*e.g.*, RT primers, target-specific primers, extension sequence primers) such that each individual molecule

produced by a primer obtains a unique barcode tag. Thus, in some embodiments, the molecular barcode tag permits a determination of whether a sequenced molecular is unique. In some embodiments, molecular barcodes may be used to silence sequencing errors, improve confidence calling of fusions or other mutations, and improved detection limits.

[0035] In some embodiments, methods are provided for preparing nucleic acids that have a target region 5' to an adjacent region (*e.g.*, an adjacent region of unknown sequence) using an oligonucleotide comprising a hairpin. For example, FIG. 3 presents a schematic of an exemplary method for amplifying target nucleic acids that have a known target region 5' to an adjacent region (for purposes of sequencing the adjacent region). At step **301**, an initial RNA is obtained or provided in a sample and is used as a template for the proceeding method. The RNA template is exposed to a plurality of hairpin primers (*e.g.*, random primers with a hairpin tail) that comprise a hairpin sequence that is 5' to different hybridization sequences and shared between all of the primers of the population. The hairpin sequence comprises two complementary common sequences that flank a molecular barcode sequence (MBC). The complementary common sequences base pair to form the stem-loop hairpin structure and protect the MBC sequence. In some embodiments, at least one primer hybridizes to the RNA molecule and primes a reverse transcriptase reaction to produce a complementary DNA strand. In step **302**, any unhybridized oligonucleotides are enzymatically degraded (*e.g.*, by an exonuclease). Also in step **302**, the RNA template is enzymatically degraded from the complementary DNA strand (*e.g.*, by RNaseH).

[0036] In step **303**, DNA molecules produced by reverse transcription are contacted by one or more initial target-specific primers which may or may not be the same as the first target-specific primer. In step **304**, hybridization of the first target-specific primer to a portion of the target nucleic acid primes an extension reaction using the DNA molecule as a template to produce a complementary DNA strand. In some embodiments, synthesis of the complementary DNA strand may reduce or eliminate hairpin formation of the complementary common sequences. Extension products are purified in step **305**.

[0037] In step **306**, DNA molecules are contacted by a first target-specific primer and a tailed primer. The first target-specific primer hybridizes to a portion of the target nucleic acid. The first tail primer hybridizes to a portion of the DNA molecule provided by the common sequence involved in hairpin formation step **301**. In some embodiments, a nested target-specific primer (*e.g.*, nested with respect to the target-specific primer of step **303**) is used in step **306**. In some embodiments, the first tailed primer may comprise an additional sequence 5' to the hybridization sequence that may include barcode, index, adapter sequences, or sequencing primer sites, for example. In step **307**, hybridization of each of the first target-specific primer and the tailed

primer allows for amplification of a portion of the target nucleic acid molecule in a polymerase chain reaction (PCR). In some embodiments, amplified products are purified in step **308**.

[0038] In step **309**, amplified DNA products (*e.g.*, those purified in step **308**) are contacted with a second target-specific primer and a common sequence primer. In some embodiments, the second target-specific primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first target-specific primer such that the reactions are nested. In some embodiments, the common sequence primer hybridizes with a sequence provided by the first tail primer in step **306**. In some embodiments, in step **310**, DNA products purified in step **308** are amplified by PCR in which the extensions are primed by a second target-specific primer and a common sequence primer. In some embodiments, amplified product from step **308** may be amplified.

[0039] In some embodiments, the second target-specific primer comprises a nucleotide sequence 5' to the target-specific sequence that comprises a barcode, index, or adapter sequences. In some embodiments, the second tail primer hybridizes to a sequence that is present within the template DNA molecule 3' of the sequence of the first tail primer such that the reactions are nested. In such embodiments, a portion of the product from step **308** is amplified. In some embodiments, the common sequence primer may comprise additional sequences 5' to the hybridization sequence that may include barcode, index, adapter sequences or sequencing primer sites. Hybridization of the second target-specific primer and the common sequence primer allows for exponential amplification of a portion of the target nucleic acid molecule in a PCR reaction. In some embodiments, products are purified in reaction **311** useful for analysis. For example, products purified in step **311** can be sequenced (*e.g.*, using a next generation sequencing platform.)

[0040] In some embodiments, all of the components involved in steps **301-311** are present at the outset and throughout the reaction. In some embodiments, steps **301-304**, **306-307**, and **309-310** are performed consecutively in a single reaction tube without any intervening purification steps. In some embodiments, steps **301-304** are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps **301-304** are present at the outset and throughout the reaction. In some embodiments, steps **306-307** are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps **306-307** are present at the outset and throughout the reaction. In some embodiments, steps **309-310** are performed consecutively in a single reaction tube. In some embodiments, all of the components involved in steps **309-310** are present at the outset and throughout the reaction.

[0041] In some embodiments, methods are provided herein that involve determining the nucleotide sequence contiguous to (adjacent to) a known target nucleotide sequence. In some

embodiments, methods comprise contacting a target nucleic acid molecule comprising the known target nucleotide sequence with an initial target-specific primer under suitable hybridization conditions. In some embodiments, the methods further comprise maintaining the target nucleic acid molecule under conditions that promote extension of the hybridized initial target-specific primer (*e.g.*, using the target nucleic acid molecule as a template), thereby producing a first extension product. In some embodiments, the methods further comprise contacting the extension product with a population of tailed random primers under suitable hybridization conditions. In some embodiments, the methods further comprise maintaining the extension product under conditions that promote extension of a hybridized tailed random primer using the portion of the target nucleic acid molecule downstream of the site of hybridization as a template, thereby producing a second extension product. In some embodiments, the methods further comprise amplifying a portion of the target nucleic acid molecule and the tailed random primer sequence with a first tail primer and a first target-specific primer, thereby producing a first amplicon. In some embodiments, the methods further comprise amplifying a portion of the amplicon with a second tail primer and a second target-specific primer, thereby producing a second amplicon.

[0042] In some embodiments, one or more target-specific primers used in the methods may be nested with respect to one or more other target-specific primers. For example, in some embodiments, a second target-specific primer is internal to a first target-specific primer. In some embodiments, target-specific primers are the same. In some embodiments, target-specific primers are nested but overlapping with respect to target complementarity. In some embodiments, target-specific primers are nested and non-overlapping. In some embodiments, combinations of identical and nested target specific primers are used in the same or different amplification steps. In some embodiments, nesting of primers increases target specificity. In some embodiments, the methods further comprise sequencing the second amplicon using a first and second sequencing primer. In some embodiments, the population of tailed random primers comprises single-stranded oligonucleotide molecules having a 5' nucleotide sequence identical to a first sequencing primer and a 3' nucleotide comprising from random nucleotides (*e.g.*, about 6 to about 12 random nucleotides). In some embodiments, the first target-specific primer comprises a nucleic acid sequence that can specifically anneal to the known nucleotide sequence of the target nucleic acid at an appropriate annealing temperature. In some embodiments, the second target-specific primer comprises a 3' portion comprising a nucleic acid sequence that can specifically anneal to a portion of the known target nucleotide sequence comprised by the first amplicon, and a 5' portion comprising a nucleic acid sequence that is identical to a second sequencing primer and the second target-specific primer is nested with respect to the first target-specific primer. In some embodiments, the first tail primer comprises a nucleic acid sequence

identical to the common sequence of the tail of the tailed random primer. In some embodiments, the common sequence on the tailed random primer is the exact match of the common sequence on the first tail primer.. In some embodiments, the second tail primer comprises a nucleic acid sequence identical to a portion of the first sequencing primer and is nested with respect to the first tail primer.

[0043] As used herein, the term “target nucleic acid” refers to a nucleic acid molecule of interest (*e.g.*, an nucleic acid to be analyzed). In some embodiments, a target nucleic acid comprises both a target nucleotide sequence (*e.g.*, a known or predetermined nucleotide sequence) and an adjacent nucleotide sequence which is to be determined (which may be referred to as an unknown sequence). A target nucleic acid can be of any appropriate length. In some embodiments, a target nucleic acid is double-stranded. In some embodiments, the target nucleic acid is DNA. In some embodiments, the target nucleic acid is genomic or chromosomal DNA (gDNA). In some embodiments, the target nucleic acid can be complementary DNA (cDNA). In some embodiments, the target nucleic acid is single-stranded. In some embodiments, the target nucleic acid can be RNA, *e.g.*, mRNA, rRNA, tRNA, long non-coding RNA, microRNA.

[0044] As used herein, the term “known target nucleotide sequence” refers to a portion of a target nucleic acid for which the sequence (*e.g.* the identity and order of the nucleotide bases of the nucleic acid) is known. For example, in some embodiments, a known target nucleotide sequence is a nucleotide sequence of a nucleic acid that is known or that has been determined in advance of an interrogation of an adjacent unknown sequence of the nucleic acid. A known target nucleotide sequence can be of any appropriate length.

[0045] In some embodiments, a target nucleotide sequence (*e.g.*, a known target nucleotide sequence) has a length of 10 or more nucleotides, 30 or more nucleotides, 40 or more nucleotides, 50 or more nucleotides, 100 or more nucleotides, 200 or more nucleotides, 300 or more nucleotides, 400 or more nucleotides, 500 or more nucleotides. In some embodiments, a target nucleotide sequence (*e.g.*, a known target nucleotide sequence) has a length in range of 10 to 100 nucleotides, 10 to 500 nucleotides, 10 to 1000 nucleotides, 100 to 500 nucleotides, 100 to 1000 nucleotides, 500 to 1000 nucleotides, 500 to 5000 nucleotides.

[0046] In some embodiments, methods are provided herein for determining sequences of contiguous (or adjacent) portions of a nucleic acid. As used herein, the term “nucleotide sequence contiguous to” refers to a nucleotide sequence of a nucleic acid molecule (*e.g.*, a target nucleic acid) that is immediately upstream or downstream of another nucleotide sequence (*e.g.*, a known nucleotide sequence). In some embodiments, a nucleotide sequence contiguous to a known target nucleotide sequence may be of any appropriate length. In some embodiments, a nucleotide sequence contiguous to a known target nucleotide sequence comprises 1 kb or less of

nucleotide sequence, *e.g.* 1 kb or less of nucleotide sequence, 750 bp or less of nucleotide sequence, 500 bp or less of nucleotide sequence, 400 bp or less of nucleotide sequence, 300 bp or less of nucleotide sequence, 200 bp or less of nucleotide sequence, 100 bp or less of nucleotide sequence. In some embodiments, in which a sample comprises different target nucleic acids comprising a known target nucleotide sequence (*e.g.* a cell in which a known target nucleotide sequence occurs multiple times in its genome, or on separate, non-identical chromosomes), there may be multiple sequences which comprise “a nucleotide sequence contiguous to” the known target nucleotide sequence. As used herein, the term “determining a (or the) nucleotide sequence,” refers to determining the identity and relative positions of the nucleotide bases of a nucleic acid.

[0047] In some embodiments of methods disclosed herein one or more tailed random primers are hybridized to a nucleic acid template (*e.g.*, a template comprising a strand of a target nucleic acid). In some embodiments, a target nucleic acid is present in or obtained from a sample comprising a plurality of nucleic acids, one or more of which plurality do not comprise the target nucleic acid. In some embodiments, one or more primers (*e.g.*, one or more tailed random primers) hybridize to substantially all of the nucleic acids in a sample. In some embodiments, one or more primers (*e.g.*, one or more tailed random primers) hybridize to nucleic acids that comprise a target nucleic acid and to nucleic acids that do not comprise the target nucleotide sequence.

[0048] Aspects of certain methods disclosed herein relate to contacting a nucleic acid template with a plurality of different primers that share a common sequence that is 5' (or upstream) to different hybridization sequences. In some embodiments the plurality of different primers may be referred to as a population of different primers. In some embodiments, the common sequence may be referred to as a tail, as such the primers are referred to as “tailed primers.” In some embodiments, different hybridization sequences of a population comprise nucleotide sequences that occur randomly or pseudorandomly within the population. In some embodiments, nucleotide sequences that occur randomly within a population contain no recognizable regularities, such that, for each nucleotide of each sequence in the population, there is an equal likelihood that the nucleotide comprises a base that is complementary with A, T, G, or C. In such embodiments, it should be appreciated that each nucleotide comprising a base that is complementary with A, T, G, or C may be a naturally occurring nucleotide, a non-naturally occurring nucleotide or a modified nucleotide.

[0049] As used herein, the term “tailed random primer” refers to a single-stranded nucleic acid molecule having a 5' nucleotide sequence (*e.g.*, a 5' nucleotide sequence identical or complementary to a first sequencing primer) and a 3' nucleic acid sequence, in which the 3'

nucleotide comprises random nucleotides (*e.g.*, from about 3 to about 15 random nucleotides, about 6 to about 12 random nucleotides). In some embodiments, the 3' nucleotide sequence comprising random nucleotides is at least 6 nucleotides in length, *e.g.* 6 nucleotides or more, 7 nucleotides or more, 8 nucleotides or more, 9 nucleotides or more, 10 nucleotides or more, 11 nucleotides or more, 12 nucleotides or more, 13 nucleotides or more, 14 nucleotides or more, 15 nucleotides or more, 20 nucleotides or more, 25 nucleotides or more in length. In some embodiments, the 3' nucleotide sequence comprising random nucleotides is 3 to 6 nucleotides in length, 3 to 9 nucleotides in length, 3 to 12 nucleotides in length, 5 to 9 nucleotides in length, 6 to 12 nucleotides in length, 3 to 25 nucleotides in length, 6 to 15 nucleotides in length, or 6 to 25 nucleotides in length. In some embodiments, a tailed random primer can further comprise a spacer between the 5' nucleotide sequence and the 3' nucleotide sequence comprising about 6 to about 12 random nucleotides. In some embodiments, the spacer is a molecular barcode, *e.g.*, that independently tags a template nucleic acid (*e.g.*, a template RNA). In some embodiments, the spacer may be 3 to 6 nucleotides in length, 3 to 12 nucleotides in length, 3 to 25 nucleotides in length, 3 to 45 nucleotides in length, 6 to 12 nucleotides in length, 8 to 16 nucleotides in length, 6 to 25 nucleotides in length, or 6 to 45 nucleotides in length. In some embodiments, for a populations of primers, the spacer is composed of random nucleotides (*e.g.*, NNNNNNNNNN, in which each of N is independently selected from A, G, C, and T). In some embodiments, the spacer (*e.g.*, a molecular barcode(MBC)) is flanked by two common regions that are complementary. In some embodiments, the complementary common regions base pair to form the stem of a hairpin having a loop portion that comprises the MBC (*e.g.*, as depicted in FIG. 3). In some embodiments, this hairpin configuration protects the MBC from annealing to other targets inhibiting the RT reaction of the extension reaction in the case of 5' fusions. In some embodiments, a population of tailed random primers can comprise individual primers with varying 3' sequences. In some embodiments, a population of tailed random primers can comprise individual primers with identical 5' nucleotide sequences, *e.g.*, they are all compatible with the same sequencing primer. In some embodiments, a population of tailed random primers can comprise individual primers with varying 5' nucleotide sequences, *e.g.* an first individual primer is compatible with a first sequencing primer and a second individual primer is compatible with a second sequencing primer.

[0050] In some embodiments, methods described herein comprises an extension regimen or step. In such embodiments, extension may proceed from one or more hybridized tailed random primers, using the nucleic acid molecules which the primers are hybridized to as templates. Extension steps are described herein. In some embodiments, one or more tailed random primers

can hybridize to substantially all of the nucleic acids in a sample, many of which may not comprise a known target nucleotide sequence. Accordingly, in some embodiments, extension of random primers may occur due to hybridization with templates that do not comprise a known target nucleotide sequence.

[0051] In some embodiments, methods described herein may involve a polymerase chain reaction (PCR) amplification regimen, involving one or more amplification cycles. As used herein, the term "amplification regimen" refers to a process of specifically amplifying (increasing the abundance of) a nucleic acid of interest. In some embodiments, exponential amplification occurs when products of a previous polymerase extension serve as templates for successive rounds of extension. In some embodiments, a PCR amplification regimen according to methods disclosed herein may comprise at least one, and in some cases at least 5 or more iterative cycles. In some embodiments each iterative cycle comprises steps of: 1) strand separation (*e.g.*, thermal denaturation); 2) oligonucleotide primer annealing to template molecules; and 3) nucleic acid polymerase extension of the annealed primers. It should be appreciated that any suitable conditions and times involved in each of these steps may be used. In some embodiments, conditions and times selected may depend on the length, sequence content, melting temperature, secondary structural features, or other factors relating to the nucleic acid template and/or primers used in the reaction. In some embodiments, an amplification regimen according to methods described herein is performed in a thermal cycler, many of which are commercially available.

[0052] In some embodiments, a nucleic acid extension reaction involves the use of a nucleic acid polymerase. As used herein, the phrase "nucleic acid polymerase" refers to an enzyme that catalyzes the template-dependent polymerization of nucleoside triphosphates to form primer extension products that are complementary to the template nucleic acid sequence. A nucleic acid polymerase enzyme initiates synthesis at the 3' end of an annealed primer and proceeds in the direction toward the 5' end of the template. Numerous nucleic acid polymerases are known in the art and commercially available. One group of nucleic acid polymerases are thermostable, *i.e.*, they retain function after being subjected to temperatures sufficient to denature annealed strands of complementary nucleic acids, *e.g.* 94 °C, or sometimes higher. A non-limiting example of a protocol for amplification involves using a polymerase (*e.g.*, VeraSeq) under the following conditions: 98 °C for 30s, followed by 14-22 cycles comprising melting at 98 °C for 10s, followed by annealing at 68 °C for 30s, followed by extension at 72 °C 3 min, followed by holding of the reaction at 4 °C. However, other appropriate reaction conditions may be used. In some embodiments, annealing/extension temperatures may be adjusted to account for differences in salt concentration (*e.g.*, 3 °C higher to higher salt concentrations).

[0053] In some embodiments, a nucleic acid polymerase is used under conditions in which the enzyme performs a template-dependent extension. In some embodiments, the nucleic acid polymerase is DNA polymerase I, Taq polymerase, Pheonix Taq polymerase, Phusion polymerase, T4 polymerase, T7 polymerase, Klenow fragment, Klenow exo-, phi29 polymerase, AMV reverse transcriptase, M-MuLV reverse transcriptase, HIV-1 reverse transcriptase, VeraSeq Ultra polymerase, VeraSeq HF 2.0 polymerase, EnzScript or another appropriate polymerase. In some embodiments, a nucleic acid polymerase is not a reverse transcriptase. In some embodiments, a nucleic acid polymerase acts on a DNA template. In some embodiments, the nucleic acid polymerase acts on an RNA template. In some embodiments, an extension reaction involves reverse transcription performed on an RNA to produce a complementary DNA molecule (RNA-dependent DNA polymerase activity). In some embodiments, a reverse transcriptase is a mouse molony murine leukemia virus (M-MLV) polymerase, AMV reverse transcriptase, RSV reverse transcriptase, HIV-1 reverse transcriptase, HIV-2 reverse transcriptase or another appropriate reverse transcriptase.

[0054] In some embodiments, a nucleic acid amplification reaction involves cycles including a strand separation step generally involving heating of the reaction mixture. As used herein, the term "strand separation" or "separating the strands" means treatment of a nucleic acid sample such that complementary double-stranded molecules are separated into two single strands available for annealing to an oligonucleotide primer. In some embodiments, strand separation according to the methods described herein is achieved by heating the nucleic acid sample above its melting temperature (T_m). In some embodiments, for a sample containing nucleic acid molecules in a reaction preparation suitable for a nucleic acid polymerase, heating to 94° C is sufficient to achieve strand separation. In some embodiments, a suitable reaction preparation contains one or more salts (e.g., 1 to 100 mM KCl, 0.1 to 10 $MgCl_2$), at least one buffering agent (e.g., 1 to 20 mM Tris-HCL), and a carrier (e.g., 0.01 to 0.5% BSA). A non-limiting example of a suitable buffer comprises 50 mM KCl, 10 mM Tris-HCl (pH 8.8@25° C), 0.5 to 3 mM $MgCl_2$, and 0.1% BSA.

[0055] In some embodiments, a nucleic acid amplification involves annealing primers to nucleic acid templates having a strands characteristic of a target nucleic acid. In some embodiments, a strand of a target nucleic acid can serve as a template nucleic acid.

[0056] As used herein, the term "anneal" refers to the formation of one or more complementary base pairs between two nucleic acids. In some embodiments, annealing involve two complementary or substantially complementary nucleic acids strands hybridizing together. In some embodiments, in the context of an extension reaction annealing involves the hybridize of primer to a template such that a primer extension substrate for a template-dependent polymerase

enzyme is formed. In some embodiments, conditions for annealing (*e.g.*, between a primer and nucleic acid template) may vary based on the length and sequence of a primer. In some embodiments, conditions for annealing are based upon a T_m (*e.g.*, a calculated T_m) of a primer. In some embodiments, an annealing step of an extension regimen involves reducing the temperature following strand separation step to a temperature based on the T_m (*e.g.*, a calculated T_m) for a primer, for a time sufficient to permit such annealing. In some embodiments, a T_m can be determined using any of a number of algorithms (*e.g.*, OLIGO™ (Molecular Biology Insights Inc. Colorado) primer design software and VENTRO NTI™ (Invitrogen, Inc. California) primer design software and programs available on the internet, including Primer3, Oligo Calculator, and NetPrimer (Premier Biosoft; Palo Alto, CA; and freely available on the world wide web (*e.g.*, at premierbiosoft.com/netprimer/netprlaunch/Help/xnetprlaunch.html)). In some embodiments, the T_m of a primer can be calculated using following formula, which is used by NetPrimer software and is described in more detail in Frieir et al. PNAS 1986 83:9373-9377 which is incorporated by reference herein in its entirety.

$$T_m = \Delta H / (\Delta S + R * \ln(C/4)) + 16.6 \log ([K^+]/(1 + 0.7 [K^+])) - 273.15$$

wherein, ΔH is enthalpy for helix formation; ΔS is entropy for helix formation; R is molar gas constant (1.987 cal/°C * mol); C is the nucleic acid concentration; and $[K^+]$ is salt concentration. For most amplification regimens, the annealing temperature is selected to be about 5° C below the predicted T_m , although temperatures closer to and above the T_m (*e.g.*, between 1° C and 5° C below the predicted T_m or between 1° C and 5° C above the predicted T_m) can be used, as can, for example, temperatures more than 5° C below the predicted T_m (*e.g.*, 6° C below, 8° C below, 10° C below or lower). In some embodiments, the closer an annealing temperature is to the T_m , the more specific is the annealing. In some embodiments, the time used for primer annealing during an extension reaction (*e.g.*, within the context of a PCR amplification regimen) is determined based, at least in part, upon the volume of the reaction (*e.g.*, with larger volumes involving longer times). In some embodiments, the time used for primer annealing during an extension reaction (*e.g.*, within the context of a PCR amplification regimen) is determined based, at least in part, upon primer and template concentrations (*e.g.*, with higher relative concentrations of primer to template involving less time than lower relative concentrations). In some embodiments, depending upon volume and relative primer/template concentration, primer annealing steps in an extension reaction (*e.g.*, within the context of an amplification regimen) can be in the range of 1 second to 5 minutes, 10 seconds and 2 minutes, or 30 seconds to 2 minutes. As used herein, “substantially anneal” refers to an extent to which complementary base pairs form between two

nucleic acids that, when used in the context of a PCR amplification regimen, is sufficient to produce a detectable level of a specifically amplified product.

[0057] As used herein, the term "polymerase extension" refers to template-dependent addition of at least one complementary nucleotide, by a nucleic acid polymerase, to the 3' end of an primer that is anneal to a nucleic acid template. In some embodiments, polymerase extension adds more than one nucleotide, *e.g.*, up to and including nucleotides corresponding to the full length of the template. In some embodiments, conditions for polymerase extension are based, at least in part, on the identity of the polymerase used. In some embodiments, the temperature used for polymerase extension is based upon the known activity properties of the enzyme. In some embodiments, in which annealing temperatures are below the optimal temperatures for the enzyme, it may be acceptable to use a lower extension temperature. In some embodiments, enzymes may retain at least partial activity below their optimal extension temperatures. In some embodiments, a polymerase extension (*e.g.*, performed thermostable polymerases) (*e.g.*, Taq polymerase and variants thereof) is performed at 65° C to 75° C or 68° C to 72° C. In some embodiments, methods provided herein involve polymerase extension of primers that are anneal to nucleic acid templates at each cycle of a PCR amplification regimen. In some embodiments, a polymerase extension is performed using a polymerase that has relatively strong strand displacement activity. In some embodiments, polymerases having strong strand displacement are useful for preparing nucleic acids for purposes of detecting fusions (*e.g.*, 5' fusions).

[0058] In some embodiments, primer extension is performed under conditions that permit the extension of annealed oligonucleotide primers. As used herein, the term "conditions that permit the extension of an annealed oligonucleotide such that extension products are generated" refers to the set of conditions including, for example temperature, salt and co-factor concentrations, pH, and enzyme concentration under which a nucleic acid polymerase catalyzes primer extension. In some embodiments, such conditions are based, at least in part, on the nucleic acid polymerase being used. In some embodiments, a polymerase may perform a primer extension reaction in a suitable reaction preparation. In some embodiments, a suitable reaction preparation contains one or more salts (*e.g.*, 1 to 100 mM KCl, 0.1 to 10 MgCl₂), at least one buffering agent (*e.g.*, 1 to 20 mM Tris-HCL), a carrier (*e.g.*, 0.01 to 0.5% BSA) and one or more NTPs (*e.g.*, 10 to 200 uM of each of dATP, dTTP, dCTP, and dGTP). non-limiting set of conditions is 50 mM KCl, 10 mM Tris-HCl (pH 8.8@25° C), 0.5 to 3 mM MgCl₂, 200 uM each dNTP, and 0.1% BSA at 72° C, under which a polymerase (*e.g.*, Taq polymerase) catalyzes primer extension. In some embodiments, conditions for initiation and extension may include the presence of one, two, three or four different deoxyribonucleoside triphosphates (*e.g.*, selected from dATP, dTTP, dCTP, and dGTP) and a polymerization-inducing agent such as DNA polymerase or reverse transcriptase, in

a suitable buffer. In some embodiments, a "buffer" may include solvents (*e.g.*, aqueous solvents) plus appropriate cofactors and reagents which affect pH, ionic strength, etc.).

[0059] In some embodiments, nucleic acid amplification involve up to 5, up to 10, up to 20, up to 30, up to 40 or more rounds (cycles) of amplification. In some embodiments, nucleic acid amplification may comprise a set of cycles of a PCR amplification regimen from 5 cycles to 20 cycles in length. In some embodiments, an amplification step may comprise a set of cycles of a PCR amplification regimen from 10 cycles to 20 cycles in length. In some embodiments, each amplification step can comprise a set of cycles of a PCR amplification regimen from 12 cycles to 16 cycles in length. In some embodiments, an annealing temperature can be less than 70 °C. In some embodiments, an annealing temperature can be less than 72 °C. In some embodiments, an annealing temperature can be about 65 °C. In some embodiments, an annealing temperature can be from about 61 to about 72 °C.

[0060] In various embodiments, methods and compositions described herein relate to performing a PCR amplification regimen with one or more of the types of primers described herein. As used herein, "primer" refers to an oligonucleotide capable of specifically annealing to a nucleic acid template and providing a 3' end that serves as a substrate for a template-dependent polymerase to produce an extension product which is complementary to the template. In some embodiments, a primer useful in methods described herein is single-stranded, such that the primer and its complement can anneal to form two strands. Primers according to methods and compositions described herein may comprise a hybridization sequence (*e.g.*, a sequence that anneals with a nucleic acid template) that is less than or equal to 300 nucleotides in length, *e.g.*, less than or equal to 300, or 250, or 200, or 150, or 100, or 90, or 80, or 70, or 60, or 50, or 40, or 30 or fewer, or 20 or fewer, or 15 or fewer, but at least 6 nucleotides in length. In some embodiments, a hybridization sequence sequence of a primer may be 6 to 50 nucleotides in length, 6 to 35 nucleotides in length, 6 to 20 nucleotides in length, 10 to 25 nucleotides in length.

[0061] Any suitable method may be used for synthesizing oligonucleotides and primers. In some embodiments, commercial sources offer oligonucleotide synthesis services suitable for providing primers for use in methods and compositions described herein, *e.g.* INVITROGEN™ Custom DNA Oligos; Life Technologies; Grand Island, NY or custom DNA Oligos from IDT; Coralville, IA).

[0062] In some embodiments, after an extension from a tailed random primer has occurred, the extension product and template can be amplified in a first amplification step. In some embodiments, amplification may involve a set of PCR amplification cycles using a first target-specific primer and a first tail primer. In some embodiments, the amplification may result in at least part of the tailed random primer sequence present in the extension product being amplified.

In some embodiments, the amplification may result in all of the tailed random primer sequence present in the extension product being amplified.

[0063] As used herein, the term “first target-specific primer” refers to a single-stranded oligonucleotide comprising a nucleic acid sequence that can specifically anneal under suitable annealing conditions to a nucleic acid template that has a strand characteristic of a target nucleic acid.

[0064] In some embodiments, a primer (*e.g.*, a target specific primer) can comprise a 5' tag sequence portion. In some embodiments, multiple primers (*e.g.*, all first-target specific primers) present in a reaction can comprise identical 5' tag sequence portions. In some embodiments, in a multiplex PCR reaction, different primer species can interact with each other in an off-target manner, leading to primer extension and subsequently amplification by DNA polymerase. In such embodiments, these primer dimers tend to be short, and their efficient amplification can overtake the reaction and dominate resulting in poor amplification of desired target sequence. Accordingly, in some embodiments, the inclusion of a 5' tag sequence in primers (*e.g.*, on target specific primer(s)) may result in formation of primer dimers that contain the same complementary tails on both ends. In some embodiments, in subsequent amplification cycles, such primer dimers would denature into single-stranded DNA primer dimers, each comprising complementary sequences on their two ends which are introduced by the 5' tag. In some embodiments, instead of primer annealing to these single stranded DNA primer dimers, an intra-molecular hairpin (a panhandle like structure) formation may occur due to the proximate accessibility of the complementary tags on the same primer dimer molecule instead of an inter-molecular interaction with new primers on separate molecules. Accordingly, in some embodiments, these primer dimers may be inefficiently amplified, such that primers are not exponentially consumed by the dimers for amplification; rather the tagged primers can remain in high and sufficient concentration for desired specific amplification of target sequences. In some embodiments, accumulation of primer dimers may be undesirable in the context of multiplex amplification because they compete for and consume other reagents in the reaction.

[0065] In some embodiments, a 5' tag sequence can be a GC-rich sequence. In some embodiments, a 5' tag sequence may comprise at least 50% GC content, at least 55% GC content, at least 60% GC content, at least 65% GC content, at least 70% GC content, at least 75% GC content, at least 80% GC content, or higher GC content. In some embodiments, a tag sequence may comprise at least 60% GC content. In some embodiments, a tag sequence may comprise at least 65% GC content.

[0066] In some embodiments, a target-specific primer (*e.g.*, a second target-specific primer) is a single-stranded oligonucleotide comprising a 3' portion comprising a nucleic acid sequence

that can specifically anneal to a portion of a known target nucleotide sequence of an amplicon of an amplification reaction, and a 5' portion comprising a tag sequence (*e.g.*, a nucleotide sequence that is identical to or complementary to a sequencing primer (*e.g.*, a second sequencing primer).

[0067] In some embodiments, a second target-specific primer of an amplification regimen is nested with respect to a first target-specific primer of the amplification regimen. In some embodiments, the second target-specific primer is nested with respect to the first target-specific primer by at least 3 nucleotides, *e.g.* by 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, or 15 or more nucleotides. In some embodiments, all of the target-specific primers (*e.g.*, second target-specific primers) used in an amplification regimen comprise the same 5' portion. In some embodiments, the 5' portion target-specific primer can be configured to suppress primer dimers as described herein.

[0068] In some embodiments, first and second target-specific primers are used in an amplification regimen that are substantially complementary to the same strand of a target nucleic acid. In some embodiments, portions of the first and second target-specific primers that specifically anneal to a target sequence (*e.g.*, a known target sequence) can comprise a total of at least 20 unique bases of the known target nucleotide sequence, *e.g.* 20 or more unique bases, 25 or more unique bases, 30 or more unique bases, 35 or more unique bases, 40 or more unique bases, or 50 or more unique bases. In some embodiments, portions of first and second target-specific primers that specifically anneal to a target sequence (*e.g.*, a known target sequence) can comprise a total of at least 30 unique bases of the known target nucleotide sequence.

[0069] As used herein, the term "first tail primer" refers to a nucleic acid molecule comprising a nucleic acid sequence identical to the tail portion of tailed primer.

[0070] As used herein, the term "second tail primer" refers to a nucleic acid molecule comprising a nucleic acid sequence identical to a portion of a first sequencing primer, adapter, index primer, *etc.* and is optionally nested with respect to a first tailed primer. In some embodiments, the second tail primer sits outside of the first tail primer to facilitate addition of appropriate index tags, adapters (*e.g.*, for use in a sequencing platform), *etc.* In some embodiments, a second tailed primer is identical to a sequencing primer. In some embodiments, a second tailed primer is complementary to a sequencing primer.

[0071] In some embodiments, a second tail primer is nested with respect to a first tail primer. In some embodiments, a second tail primer is not nested with respect to a first tail primer. In some embodiments, tail primers of an amplification regimen are nested with respect to one another by at least 3 nucleotides, *e.g.* by 3 nucleotides, by 4 nucleotides, by 5 nucleotides, by 6 nucleotides, by 7 nucleotides, by 8 nucleotides, by 9 nucleotides, by 10 nucleotides or more.

[0072] In some embodiments, a first tail primer comprises a nucleic acid sequence identical to or complementary to the extension product of step (b) strand which is not comprised by the second tail primer and which is located closer to the 5' end of the tailed random primer than any of the sequence identical to or complementary to the second tail primer. Thus, in some embodiments, a second tail primer sits outside of a region added by a random tail primer (5' end), *e.g.*, within the 5' tail added by the first tail primers.

[0073] In some embodiments, a first tail primer can comprise a nucleic acid sequence identical to or complementary to a stretch (*e.g.*, of about 20 nucleotides) of the 5'-most nucleotides of a tailed random primer, and a second tail primer can comprise a nucleic acid sequence identical to or complementary to about 30 bases of a tailed random primer, with a 5' nucleotides that is at least 3 nucleotides 3' of the 5' terminus of the tailed random primer.

[0074] In some embodiments, use of nested tail primers minimizes or eliminates the production of final amplicons that are amplifiable (*e.g.* during bridge PCR or emulsion PCR) but cannot be sequenced, a situation that can arise during hemi-nested methods. In some embodiments, hemi-nested approaches using a primer identical to a sequencing primer can result in the carry-over of undesired amplification products from a first PCR step to a second PCR step and may yield artificial sequencing reads. In some embodiments, the use of two tail primers, as described herein can reduce, and in some embodiments eliminate, these problems.

[0075] In some embodiments, in a first PCR amplification cycle of a first amplification step, a first target-specific primer can specifically anneal to a template strand of any nucleic acid comprising the known target nucleotide sequence. In some embodiments, depending upon the orientation with which the first target-specific primer was designed, sequence upstream or downstream of the known target nucleotide sequence, and complementary to the template strand will be synthesized. In some embodiments, in which an extension product is formed that comprises the hybridization sequence with which the first target-specific primer forms complementary base pairs, a double-stranded amplification product can be formed that comprises the first target-specific primer (and the sequence complementary thereto), the target nucleotide sequence downstream of the first target-specific primer (and the sequence complementary thereto), and the tailed random primer sequence (and the sequence complementary thereto). In such embodiments, in subsequent PCR amplification cycles, both the first target-specific primer and the first tail primer are capable of specifically annealing to appropriate strands of the amplification product and the sequence between the known nucleotide target sequence and the tailed random primer can be amplified.

[0076] In some embodiments, of methods described herein, a portion of an amplified product (an amplicon) is amplified in further rounds of amplification. In some embodiments, the further

rounds of amplification may involve PCR amplification cycles performed using a second target-specific primer and a first sequencing primer or a second tail primer. In some embodiments, a PCR amplification cycles may involve the use of PCR parameters identical to, or which differ from, those of one or more other (*e.g.*, prior) of PCR amplification cycles. In some embodiments, PCR amplification regimens can have the same or different annealing temperatures or the same or different extension step time lengths.

[0077] In some embodiments, methods described herein allow for determining the nucleotide sequence contiguous to a known target nucleotide sequence on either or both flanking regions of the known target nucleotide sequence. Regardless of whether the target nucleic acid normally exists as a single-stranded or double-stranded nucleic acid, sequence information may be represented in a single-stranded format (Strand A), from 5' to 3'. In some embodiments, if the sequence 5' to a known target nucleotide sequence of Strand A is to be determined, gene-specific primers can be complementary to (anneal to) Strand A. If the sequence 3' to a known target nucleotide sequence of Strand A is to be determined, the gene-specific primers can be identical to Strand A, such that they will anneal to the complementary strand of a double-stranded target nucleic acid.

[0078] In some embodiments, methods described herein, relating to the use of a first and second gene-specific primer can result in assays with a superior on-target rate, *e.g.* 70-90%. In some embodiments, the assays and methods described herein can have a target specificity rate of at least 85%.

[0079] In some embodiments, primers disclosed herein (*e.g.*, target-specific primers, tail primers) are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of from about 61 to 72°C, *e.g.* from about 61 to 69 °C, from about 63 to 69 °C, from about 63 to 67 °C, from about 64 to 66 °C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of less than 72 °C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of less than 70 °C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of less than 68 °C. In some embodiments, primers disclosed herein are designed such that they will specifically anneal to their complementary sequences at an annealing temperature of about 65 °C.

[0080] In some embodiments, portions of the target-specific primers that specifically anneal to the known target nucleotide sequence will anneal specifically at a temperature of about 61 to 72°C, *e.g.* from about 61 to 69 °C, from about 63 to 69 °C, from about 63 to 67 °C, from about 64

to 66 °C. In some embodiments, portions of the target-specific primers that specifically anneal to the known target nucleotide sequence will anneal specifically at a temperature of about 65°C in a PCR buffer.

[0081] In some embodiments, primers described herein do not comprise modified bases (*e.g.* the primers can not comprise a blocking 3' amine). However, in some embodiments, primers described herein do comprise modified or non-naturally occurring bases. In some embodiments, primers may be modified with a label capable of providing a detectable signal, either directly or indirectly. Non-limiting examples of such labels include radioisotopes, fluorescent molecules, biotin, and others. In some embodiments, primers disclosed herein may include contain a biotin linker or other suitable linker (*e.g.*, for conjugating the primer to a support). In some embodiments, primer may contain a target sequence of an endonucleases such that cleavage with the appropriate enzyme. In other embodiments, the 5' end of a primer may include a sequence that is complementary with a nucleic acid bound to a bead or other support, *e.g.*, a flow cell substrate. Primers may or may not comprise modified internucleoside linkages.

[0082] In some embodiments, of methods described herein, nucleic acids (*e.g.*, amplified nucleic acids, extension products, target nucleic acids) can be sequenced. In some embodiments, sequencing can be performed by a next-generation sequencing method. As used herein “next-generation sequencing” refers to oligonucleotide sequencing technologies that have the capacity to sequence oligonucleotides at speeds above those possible with conventional sequencing methods (*e.g.* Sanger sequencing), due to performing and reading out thousands to millions of sequencing reactions in parallel. Non-limiting examples of next-generation sequencing methods/platforms include Massively Parallel Signature Sequencing (Lynx Therapeutics); 454 pyro-sequencing (454 Life Sciences/ Roche Diagnostics); solid-phase, reversible dye-terminator sequencing (Solexa/Illumina); SOLiD technology (Applied Biosystems); Ion semiconductor sequencing (ION Torrent); DNA nanoball sequencing (Complete Genomics); and technologies available from Pacific Biosciences, Intelligen Bio-systems, Oxford Nanopore Technologies, and Helicos Biosciences. In some embodiments, the sequencing primers can comprise portions compatible with the selected next-generation sequencing method. Next-generation sequencing technologies and the constraints and design parameters of associated sequencing primers are well known in the art (see, *e.g.* Shendure, et al., “Next-generation DNA sequencing,” *Nature*, 2008, vol. 26, No. 10, 1135-1145; Mardis, “The impact of next-generation sequencing technology on genetics,” *Trends in Genetics*, 2007, vol. 24, No. 3, pp. 133-141; Su, et al., “Next-generation sequencing and its applications in molecular diagnostics” *Expert Rev Mol Diagn*, 2011, 11(3):333-43; Zhang et al., “The impact of next-generation sequencing on genomics”, *J Genet Genomics*, 2011, 38(3):95-109; (Nyren, P. et al. *Anal Biochem* 208: 17175 (1993); Bentley, D. R.

Curr Opin Genet Dev 16:545-52 (2006); Strausberg, R. L., et al. Drug Disc Today 13:569-77 (2008); U.S. Pat. No. 7,282,337; U.S. Pat. No. 7,279,563; U.S. Pat. No. 7,226,720; U.S. Pat. No. 7,220,549; U.S. Pat. No. 7,169,560; U.S. Pat. No. 6,818,395; U.S. Pat. No. 6,911,345; US Pub. Nos. 2006/0252077; 2007/0070349; and 20070070349; which are incorporated by reference herein in their entirety).

[0083] In some embodiments, the sequencing step involve the use of a first and second sequencing primers. In some embodiments, the first and second sequencing primers are selected to be compatible with a next-generation sequencing method as described herein.

[0084] Methods of aligning sequencing reads to known sequence databases of genomic and/or cDNA sequences are well known in the art and software is commercially available for this process. In some embodiments, reads (less the sequencing primer nucleotide sequence) which do not map, in their entirety, to wild-type sequence databases can be genomic rearrangements or large indel mutations. In some embodiments, reads (less the sequencing primer nucleotide sequence) comprising sequences which map to multiple locations in the genome can be genomic rearrangements.

[0085] In some embodiments, primers may contain additional sequences such as an identifier sequence (*e.g.*, a barcode, an index), sequencing primer hybridization sequences (*e.g.*, Rd1), and adapter sequences. In some embodiments the adapter sequences are sequences used with a next generation sequencing system. In some embodiments, the adapter sequences are P5 and P7 sequences for Illumina-based sequencing technology. In some embodiments, the adapter sequence are P1 and A compatible with Ion Torrent sequencing technology.

[0086] In some embodiments, barcode or index sequences may provide a source or location identifier for a target nucleic acid. For example, a barcode or index sequence may serve to identify a patient from whom a nucleic acid is obtained. In some embodiments, barcode or index sequences enable sequencing of multiple different samples on a single reaction (*e.g.*, performed in a single flow cell). In some embodiments, an index sequence can be used to orientate a sequence imager for purposes of detecting individual sequencing reactions. In some embodiments, a barcode or index sequence may be 2 to 25 nucleotides in length, 2 to 15 nucleotides in length, 2 to 10 nucleotides in length, 2 to 6 nucleotides in length.

[0087] In some embodiments, when a population of tailed random primers is used in accordance with methods described herein, multiple distinguishable amplification products can be present after amplification. In some embodiments, because tailed random primers hybridize at various positions throughout nucleic acid molecules of a sample, a set of target-specific primers can hybridize (and amplify) the extension products created by more than 1 hybridization event, *e.g.* one tailed random primer may hybridize at a first distance (*e.g.*, 100 nucleotides) from a

target-specific primer hybridization site, and another tailed random primer can hybridize at a second distance (*e.g.*, 200 nucleotides) from a target-specific primer hybridization site, thereby resulting in two amplification products (*e.g.*, a ~100 bp amplification product and a ~200 bp amplification product). In some embodiments, these multiple amplification products can each be sequenced in . In some embodiments, sequencing of these multiple amplification products is advantageous because it provides multiple overlapping sequence reads that can be compare with one another to detect sequence errors introduced during amplification or sequencing processes. In some embodiments, individual amplification products can be aligned and where they differ in the sequence present at a particular base, an artifact or error of PCR and/or sequencing may be present.

[0088] In some embodiments, target nucleic acids and/or amplification products thereof can be isolated from enzymes, primers, or buffer components before and/or after any of appropriate step of a method. Any suitable methods for isolating nucleic acids may be used. In some embodiments, the isolation can comprise Solid Phase Reversible Immobilization (SPRI) cleanup. Methods for SPRI cleanup are well known in the art and kits are commercially available, *e.g.* Agencourt AMPure XP - PCR Purification (Cat No. A63880, Beckman Coulter; Brea, CA). In some embodiments, enzymes can be inactivated by heat treatment.

[0089] In some embodiments, unhybridized primers can be removed from a nucleic acid preparation using appropriate methods (*e.g.*, purification, digestion, *etc.*). In some embodiments, a nuclease (*e.g.*, exonuclease I) is used to remove primer from a preparation. In some embodiments, such nucleases are heat inactivated subsequent to primer digestion. Once the nucleases are inactivated a further set of primers may be added together with other appropriate components (*e.g.*, enzymes, buffers) to perform a further amplification reaction.

[0090] In some embodiments, a target nucleic acid genomic DNA or a portion thereof. In some embodiments, a target nucleic acid can be ribonucleic acid (RNA), *e.g.* mRNA, or a portion thereof. In some embodiments, a target nucleic acid can be a cDNA or a portion thereof.

[0091] Many of the sequencing methods suitable for use in methods described herein provide sequencing runs with optimal read lengths of tens to hundreds of nucleotide bases (*e.g.* Ion Torrent technology can produce read lengths of 200-400 bp). Target nucleic acids may or may not be substantially longer than this optimal read length. In some embodiments, in order for an amplified nucleic acid portion to be of a suitable length for use in a particular sequencing technology, the average distance between the known target nucleotide sequence and an end of the target nucleic acid to which a tailed random primer is hybridizable should be as close to the optimal read length of the selected technology as possible. In some embodiments, if the optimal read-length of a given sequencing technology is 200 bp, then the nucleic acid molecules

amplified in accordance with methods described herein should have an average length of about 800 bp, about 700 bp, about 600 bp, about 500 bp, about 400 bp, about 300 bp, about 200 bp or less.

[0092] Nucleic acids used herein (*e.g.*, prior to sequencing) can be sheared, *e.g.* mechanically or enzymatically sheared, to generate fragments of any desired size. Non-limiting examples of mechanical shearing processes include sonication, nebulization, and AFA™ shearing technology available from Covaris (Woburn, MA). In some embodiments, a nucleic acid can be mechanically sheared by sonication.

[0093] In some embodiments, a target nucleic acid is not sheared or digested. In some embodiments, nucleic acid products of preparative steps (*e.g.*, extension products, amplification products) are not sheared or enzymatically digested.

[0094] In some embodiments, when a target nucleic acid is an RNA, the sample can be subjected to a reverse transcriptase regimen to generate a DNA template and the DNA template can then be sheared. In some embodiments, target RNA can be sheared before performing a reverse transcriptase regimen. In some embodiments, a sample comprising target RNA can be used in methods described herein using total nucleic acids extracted from either fresh or degraded specimens; without the need of genomic DNA removal for cDNA sequencing; without the need of ribosomal RNA depletion for cDNA sequencing; without the need of mechanical or enzymatic shearing in any of the steps; by subjecting the RNA for double-stranded cDNA synthesis using random hexamers.

[0095] In some embodiments, a known target nucleic acid can contain a fusion sequence resulting from a gene rearrangement. In some embodiments, methods described herein are suited for determining the presence and/or identity of a gene rearrangement. In some embodiments, identity of one portion of a gene rearrangement is previously known (*e.g.*, the portion of a gene rearrangement that is to be targeted by the gene-specific primers) and the sequence of the other portion may be determined using methods disclosed herein. In some embodiments, a gene rearrangement can involve an oncogene. In some embodiments, a gene rearrangement can comprise a fusion oncogene.

[0096] In some embodiments, a target nucleic acid is present in or obtained from an appropriate sample (*e.g.*, a food sample, environmental sample, biological sample *e.g.*, blood sample, etc.). In some embodiments, there is a biological sample obtained from a subject. In some embodiments a sample can be a diagnostic sample obtained from a subject. In some embodiments, a sample can further comprise proteins, cells, fluids, biological fluids, preservatives, and/or other substances. By way of non-limiting example, a sample can be a cheek swab, blood, serum, plasma, sputum, cerebrospinal fluid, urine, tears, alveolar isolates, pleural

fluid, pericardial fluid, cyst fluid, tumor tissue, tissue, a biopsy, saliva, an aspirate, or combinations thereof. In some embodiments, a sample can be obtained by resection or biopsy.

[0097] In some embodiments, the sample can be obtained from a subject in need of treatment for a disease associated with a genetic alteration, *e.g.* cancer or a hereditary disease. In some embodiments, a known target sequence is present in a disease-associated gene.

[0098] In some embodiments, a sample is obtained from a subject in need of treatment for cancer. In some embodiments, the sample comprises a population of tumor cells, *e.g.* at least one tumor cell. In some embodiments, the sample comprises a tumor biopsy, including but not limited to, untreated biopsy tissue or treated biopsy tissue (*e.g.* formalin-fixed and/or paraffin-embedded biopsy tissue).

[0099] In some embodiments, the sample is freshly collected. In some embodiments, the sample is stored prior to being used in methods and compositions described herein. In some embodiments, the sample is an untreated sample. As used herein, “untreated sample” refers to a biological sample that has not had any prior sample pre-treatment except for dilution and/or suspension in a solution. In some embodiments, a sample is obtained from a subject and preserved or processed prior to being utilized in methods and compositions described herein. By way of non-limiting example, a sample can be embedded in paraffin wax, refrigerated, or frozen. A frozen sample can be thawed before determining the presence of a nucleic acid according to methods and compositions described herein. In some embodiments, the sample can be a processed or treated sample. Exemplary methods for treating or processing a sample include, but are not limited to, centrifugation, filtration, sonication, homogenization, heating, freezing and thawing, contacting with a preservative (*e.g.* anti-coagulant or nuclease inhibitor) and any combination thereof. In some embodiments, a sample can be treated with a chemical and/or biological reagent. Chemical and/or biological reagents can be employed to protect and/or maintain the stability of the sample or nucleic acid comprised by the sample during processing and/or storage. In addition, or alternatively, chemical and/or biological reagents can be employed to release nucleic acids from other components of the sample. By way of non-limiting example, a blood sample can be treated with an anti-coagulant prior to being utilized in methods and compositions described herein. Suitable methods and processes for processing, preservation, or treatment of samples for nucleic acid analysis may be used in the method disclosed herein. In some embodiments, a sample can be a clarified fluid sample, for example, by centrifugation. In some embodiments, a sample can be clarified by low-speed centrifugation (*e.g.* 3,000 x g or less) and collection of the supernatant comprising the clarified fluid sample.

[00100] In some embodiments, a nucleic acid present in a sample can be isolated, enriched, or purified prior to being utilized in methods and compositions described herein. Suitable methods

of isolating, enriching, or purifying nucleic acids from a sample may be used. For example, kits for isolation of genomic DNA from various sample types are commercially available (*e.g.*, Catalog Nos. 51104, 51304, 56504, and 56404; Qiagen; Germantown, MD). In some embodiments, methods described herein relate to methods of enriching for target nucleic acids, *e.g.*, prior to a sequencing of the target nucleic acids. In some embodiments, a sequence of one end of the target nucleic acid to be enriched is not known prior to sequencing. In some embodiments, methods described herein relate to methods of enriching specific nucleotide sequences prior to determining the nucleotide sequence using a next-generation sequencing technology. In some embodiments, methods of enriching specific nucleotide sequences do not comprise hybridization enrichment.

[00101] Methods described herein can be employed in a multiplex format. In embodiments of methods described herein, multiplex applications can include determining the nucleotide sequence contiguous to one or more known target nucleotide sequences. As used herein, “multiplex amplification” refers to a process involve simultaneous amplification of more than one target nucleic acid in one reaction vessel. In some embodiments, methods involve subsequent determination of the sequence of the multiplex amplification products using one or more sets of primers. Multiplex can refer to the detection of between about 2-1,000 different target sequences in a single reaction. As used herein, multiplex refers to the detection of any range between 2-1,000, *e.g.*, between 5-500, 25-1000, or 10-100 different target sequences in a single reaction, etc. The term “multiplex” as applied to PCR implies that there are primers specific for at least two different target sequences in the same PCR reaction.

[00102] In some embodiments, target nucleic acids in a sample, or separate portions of a sample, can be amplified with a plurality of primers (*e.g.*, a plurality of first and second target-specific primers). In some embodiments, the plurality of primers (*e.g.*, a plurality of first and second target-specific primers) can be present in a single reaction mixture, *e.g.* multiple amplification products can be produced in the same reaction mixture. In some embodiments, the plurality of primers (*e.g.*, a plurality of sets of first and second target-specific primers) can specifically anneal to known target sequences comprised by separate genes. In some embodiments, at least two sets of primers (*e.g.*, at least two sets of first and second target-specific primers) can specifically anneal to different portions of a known target sequence. In some embodiments, at least two sets of primers (*e.g.*, at least two sets of first and second target-specific primers) can specifically anneal to different portions of a known target sequence comprised by a single gene. In some embodiments, at least two sets of primers (*e.g.*, at least two sets of first and second target-specific primers) can specifically anneal to different exons of a gene comprising a

known target sequence. In some embodiments, the plurality of primers (*e.g.*, first target-specific primers) can comprise identical 5' tag sequence portions.

[00103] In embodiments of methods described herein, multiplex applications can include determining the nucleotide sequence contiguous to one or more known target nucleotide sequences in multiple samples in one sequencing reaction or sequencing run. In some embodiments, multiple samples can be of different origins, *e.g.* from different tissues and/or different subjects. In such embodiments, primers (*e.g.*, tailed random primers) can further comprise a barcode portion. In some embodiments, a primer (*e.g.*, a tailed random primer) with a unique barcode portion can be added to each sample and ligated to the nucleic acids therein; the samples can subsequently be pooled. In such embodiments, each resulting sequencing read of an amplification product will comprise a barcode that identifies the sample containing the template nucleic acid from which the amplification product is derived.

[00104] In some embodiments of methods described herein, a determination of the sequence contiguous to a known oligonucleotide target sequence can provide information relevant to treatment of disease. Thus, in some embodiments, methods disclosed herein can be used to aid in treating disease. In some embodiments, a sample can be from a subject in need of treatment for a disease associated with a genetic alteration. In some embodiments, a known target sequence a sequence of a disease-associated gene, *e.g.* an oncogene. In some embodiments, a sequence contiguous to a known oligonucleotide target sequence and/or the known oligonucleotide target sequence can comprise a mutation or genetic abnormality which is disease-associated, *e.g.* a SNP, an insertion, a deletion, and/or a gene rearrangement. In some embodiments, a sequence contiguous to a known target sequence and/or a known target sequence present in a sample comprised sequence of a gene rearrangement product. In some embodiments, a gene rearrangement can be an oncogene, *e.g.* a fusion oncogene.

[00105] Certain treatments for cancer are particularly effective against tumors comprising certain oncogenes, *e.g.* a treatment agent which targets the action or expression of a given fusion oncogene can be effective against tumors comprising that fusion oncogene but not against tumors lacking the fusion oncogene. Methods described herein can facilitate a determination of specific sequences that reveal oncogene status (*e.g.* mutations, SNPs, and/or rearrangements). In some embodiments, methods described herein can further allow the determination of specific sequences when the sequence of a flanking region is known, *e.g.* methods described herein can determine the presence and identity of gene rearrangements involving known genes (*e.g.*, oncogenes) in which the precise location and/or rearrangement partner are not known before methods described herein are performed.

[00106] In some embodiments, technology described herein relates to a method of treating cancer. Accordingly, in some embodiments, methods provided herein may involve detecting, in a tumor sample obtained from a subject in need of treatment for cancer, the presence of one or more oncogene rearrangements; and administering a cancer treatment which is effective against tumors having any of the detected oncogene rearrangements. In some embodiments, technology described herein relates to a method of determining if a subject in need of treatment for cancer will be responsive to a given treatment. Accordingly, in some embodiments, methods provided herein may involve detecting, in a tumor sample obtained from a subject, the presence of an oncogene rearrangement, in which the subject is determined to be responsive to a treatment targeting an oncogene rearrangement product if the presence of the oncogene rearrangement is detected.

[00107] In some embodiments, a subject is in need of treatment for lung cancer. In some embodiments, *e.g.* when the sample is obtained from a subject in need of treatment for lung cancer, the known target sequence can comprise a sequence from a gene selected from the group of ALK, ROS1, and RET. Accordingly, in some embodiments, gene rearrangements result in fusions involving the ALK, ROS1, or RET. Non-limiting examples of gene arrangements involving ALK, ROS1, or RET are described in, *e.g.*, Soda et al. Nature 2007 448:561-6; Rikova et al. Cell 2007 131:1190-1203; Kohno et al. Nature Medicine 2012 18:375-7; Takouchi et al. Nature Medicine 2012 18:378-81; which are incorporated by reference herein in their entireties. However, it should be appreciated that the precise location of a gene rearrangement, and the identity of the second gene involved in the rearrangement may not be known in advance. Accordingly, in methods described herein, the presence and identity of such rearrangements can be detected without having to know the location of the rearrangement or the identity of the second gene involved in the gene rearrangement.

[00108] In some embodiments, the known target sequence can comprise sequence from a gene selected from the group of: ALK, ROS1, and RET.

[00109] In some embodiments, the presence of a gene rearrangement of ALK in a sample obtained from a tumor in a subject can indicate that the tumor is susceptible to treatment with a treatment selected from the group consisting of: an ALK inhibitor; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; diamino and aminopyrimidine inhibitors of ALK kinase activity such as NVP-TAE684 and PF-02341066 (see, *e.g.* Galkin et al, Proc Natl Acad Sci USA, 2007, 104:270-275; Zou et al. Cancer Res, 2007, 67:4408-4417; Hallberg and Palmer F1000 Med Reports 2011 3:21; and Sakamoto et al. Cancer Cell 2011 19:679-690) and molecules disclosed in WO 04/079326. All of the foregoing references are incorporated by reference herein in their entireties. An ALK

inhibitor can include any agent that reduces the expression and/or kinase activity of ALK or a portion thereof, including, *e.g.* oligonucleotides, small molecules, and/or peptides that reduce the expression and/or activity of ALK or a portion thereof. As used herein “anaplastic lymphoma kinase” or “ALK” refers to a transmembrane tyROS1ne kinase typically involved in neuronal regulation in the wildtype form. The nucleotide sequence of the ALK gene and mRNA are known for a number of species, including human (*e.g.* SEQ ID NO: 2 (mRNA), NCBI Gene ID: 238).

[00110] In some embodiments, the presence of a gene rearrangement of ROS1 in a sample obtained from a tumor in a subject can indicate that the tumor is susceptible to treatment with a treatment selected from the group consisting of: a ROS1 inhibitor and an ALK inhibitor as described herein above (*e.g.* crizotinib). A ROS1 inhibitor can include any agent that reduces the expression and/or kinase activity of ROS1 or a portion thereof, including, *e.g.* oligonucleotides, small molecules, and/or peptides that reduce the expression and/or activity of ROS1 or a portion thereof. As used herein “c-ros oncogene 1” or “ROS1” (also referred to in the art as ros-1) refers to a transmembrane tyrosine kinase of the sevenless subfamily and which interacts with PTPN6. Nucleotide sequences of the ROS1 gene and mRNA are known for a number of species, including human (*e.g.* SEQ ID NO: 1 (mRNA), NCBI Gene ID: 238).

[00111] In some embodiments, the presence of a gene rearrangement of RET in a sample obtained from a tumor in a subject can indicate that the tumor is susceptible to treatment with a treatment selected from the group consisting of: a RET inhibitor; DP-2490, DP-3636, SU5416; BAY 43-9006, BAY 73-4506 (regorafenib), ZD6474, NVP-AST487, sorafenib, RPI-1, XL184, vandetanib, sunitinib, imatinib, pazopanib, axitinib, motesanib, gefitinib, and withaferin A (see, *e.g.* Samadi et al. Surgery 2010 148:1228-36; Cuccuru et al. JNCI 2004 13:1006-1014; Akeno-Stuart et al. Cancer Research 2007 67:6956; Grazma et al. J Clin Oncol 2010 28:15s 5559; Mologni et al. J Mol Endocrinol 2006 37:199-212; Calmomagno et al. Journal NCI 2006 98:326-334; Mologni. Curr Med Chem 2011 18:162-175 and the compounds disclosed in WO 06/034833; US Patent Publication 2011/0201598 and US Patent 8,067,434). All of the foregoing references are incorporated by reference herein in their entireties. A RET inhibitor can include any agent that reduces the expression and/or kinase activity of RET or a portion thereof, including, *e.g.* oligonucleotides, small molecules, and/or peptides that reduce the expression and/or activity of RET or a portion thereof. As used herein “rearranged during transfection” or “RET” refers to a receptor tyrosine kinase of the cadherein superfamily which is involved in neural crest development and recognizes glial cell line-derived neurotrophic factor family signaling molecules. Nucleotide sequences of the ROS1 gene and mRNA are known for a number of species, including human (*e.g.* SEQ ID NOs: 3-4 (mRNA), NCBI Gene ID: 5979).