

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

QIAGEN Sciences, LLC,
Petitioner,

v.

Tecan Genomics, Inc.,
Patent Owner.

Case IPR2025-00027

U.S. Patent No. 11,725,241

Declaration of Michael L. Metzker, Ph.D.

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	1
II. MATERIALS REVIEWED	1
III. SUMMARY OF OPINIONS.....	3
IV. EXPERIENCE AND QUALIFICATIONS.....	5
V. TECHNICAL BACKGROUND	12
A. The relevant art.....	12
B. DNA	13
C. PCR.....	17
D. DNA sequencing technologies.....	22
E. The use of molecular identifiers to detect PCR duplicates has been well-known in the art.....	25
VI. APPLICABLE STANDARDS.....	26
A. Relevant time period	27
B. Level of ordinary skill in the art.....	27
C. Anticipation.....	28
D. Obviousness.....	29
VII. THE '241 PATENT.....	31
VIII. CLAIM CONSTRUCTION	35
IX. PRIOR ART.....	36
A. <i>The '640 Patent Pub. (McCloskey)</i>	36
B. <i>The '202 Patent Pub. (Porreca)</i>	42
C. <i>The '389 PCT Appl. (Schmitt)</i>	47
D. <i>Li (2009)</i>	50
E. <i>Illumina (2008)</i>	52
F. <i>Metzker (2010)</i>	53
X. ASSERTED GROUNDS.....	56
A. Ground 1: Claims 1-3, 5-7, 9-10, and 13-16 Are Anticipated By the '640 Patent Pub. (McCloskey)	56

B.	Ground II: Claims 1-3, 5-11, 13-16 Are Anticipated by the '202 Patent Pub. (Porreca)	87
C.	Ground III: Claims 1-3, 5-12, and 14-15 Are Anticipated by the '389 PCT Appl. (Schmitt)	115
D.	Grounds IV and V: Claim 4 of the '241 Patent would have been obvious over the '202 Patent Pub. (Porreca) in view of Illumina (2008) and over '389 PCT Appl. (Schmitt) in view of Illumina (2008)	146
E.	Grounds VI and VII: Claim 8 would have been obvious over the '202 Patent Pub. (Porreca) in view of Illumina (2008) and over the '389 PCT Appl. (Schmitt) in view of Illumina (2008).....	149
F.	Ground VIII: Claim 12 would have been obvious over the '202 Patent Pub. (Porreca) in view of Li (2009)	152

I. INTRODUCTION

1. I, Michael L. Metzker, have been retained as an expert witness on behalf of QIAGEN Sciences, LLC (the “Petitioner”) in connection with the above-captioned Inter Partes review (“IPR”). I understand that the petition for IPR involves the ’241 Patent.¹

II. MATERIALS REVIEWED

2. I have reviewed the materials cited herein, including the following exhibits. The materials are publications that one of ordinary skill in the art² would have read and relied upon.

Ex.	Description
1001	U.S. Patent No. 11,725,241
1002	U.S. Patent No. 11,098,357
1003	Expert Declaration of Michael L. Metzker, Ph.D.
1004	Miner et al, <i>Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR</i> , NUCLEIC ACID RESEARCH 32:17 (2004) (“Miner 2004”)
1005	U.S. Patent Publication No. 2007/0020640 (“’640 Patent Pub. (McCloskey)”)
1006	U.S. Patent Publication No. 2012/0165202 (“’202 Patent Pub. (Porreca)”)
1007	WO 2013/142378 (“’389 PCT Appl. (Schmitt)”)
1008	Illumina Publication No. 770-2008-011, <i>Multiplexed sequencing with the Illumina Genome Analyzer system</i> (2008) (“Illumina (2008)”), available at

¹ See Ex. 1001, Amorese et al., *Compositions and methods for identification of a duplicate sequencing read*, U.S. Patent No. 11,725,241 (2020) (the “’241 Patent”).

² Throughout this Report, I also refer to one of ordinary skill in the art as a POSA.

Ex.	Description
	https://web.archive.org/web/20101226144356/https://www.illumina.com/documents/products/datasheets/datasheet_sequencing_multiplex.pdf
1009	Authenticating affidavit of Nathaniel Frank-White from the Internet Archive, including the archived copy of <i>Illumina 2008</i>
1010	Metzker, <i>Sequencing technologies – the next generation</i> , NATURE REVIEWS GENETICS 11:31-46 (2010) (“Metzker (2010)”)
1011	Li <i>et al.</i> , <i>The Sequence Alignment/Map format and SAMtools</i> , BIOINFORMATICS 25:2078-2079 (2009) (“Li (2009)”)
1012	Prosecution History of U.S. Patent No. 11,725,241
1013	Prosecution History of U.S. Patent No. 11,098,357
1014	Authenticating declaration of Sylvia D. Hall-Ellis, Ph.D.
1015	McCloskey <i>et al.</i> , <i>Encoding PCR products with batch-stamps and barcodes</i> , BIOCHEMICAL GENETICS 45:761-767 (2007) (“McCloskey (2007)”)
1016	Curriculum Vitae of Michael L. Metzker, Ph.D.
1017	Metzker & Caskey, <i>Polymerase chain reaction</i> , In <i>ENCYCLOPEDIA OF MEDICAL DEVICES AND INSTRUMENTATION</i> , Second Edition, Volume 5, pp. 380-387 (2006) (“Metzker (2006)”)
1018	Saiki <i>et al.</i> , <i>Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia</i> , SCIENCE 230:1350-1354 (1985) (“Saiki (1985)”)
1019	Saiki <i>et al.</i> , <i>Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase</i> , SCIENCE 239:487-491 (1988) (“Saiki (1988)”)
1020	Gibbs <i>et al.</i> , <i>Detection of single DNA base differences by competitive oligonucleotide priming</i> , NUCLEIC ACIDS RESEARCH 17:2437-2448 (1989) (“Gibbs (1989)”)
1021	Newton <i>et al.</i> , <i>Analysis of any point mutation in DNA: The amplification refractory mutation system (ARMS)</i> , NUCLEIC ACIDS RESEARCH 17:2503-2516 (1989) (“Newton (1989)”)
1022	Chamberlain <i>et al.</i> , <i>Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification</i> , NUCLEIC ACIDS

Ex.	Description
	RESEARCH 16:11141-11156 (1988) (“Chamberlain (1989)”)
1023	Gibbs <i>et al.</i> <i>Multiplex DNA deletion detection and exon sequencing of the hypoxanthine phosphoribosyltransferase gene in Lesch-Nyhan families</i> , GENOMICS 7:235-244 (1990) (“Gibbs (1990)”)
1024	Grompe, <i>The rapid detection of unknown mutations in nucleic acids</i> , NATURE GENETICS 5:111-117 (1993) (“Grompe (1993)”)
1025	Syvänen, <i>Accessing genetic variation: genotyping single nucleotide polymorphisms</i> , NATURE GENETICS 2:930-942 (2011) (“Syvänen (2001)”)
1026	Metzker, <i>Emerging technologies in DNA sequencing</i> , GENOME RESEARCH 15:1767-1776 (2005) (“Metzker (2005)”)
1027	Rothberg <i>et al.</i> , <i>An integrated semiconductor device enabling non-optical genome sequencing</i> , NATURE 475:348-352 (2011) (“Rothberg (2011)”)
1028	Jain <i>et al.</i> , <i>Improved data analysis for the MinION nanopore sequencer</i> , NATURE METHODS 12:351-356 (2015) (“Jain (2015)”)
1029	Illumina, <i>Multiplexing, Sample preparation guide</i> (2010) (“Illumina Manual (2010)”)
1030	Illumina, <i>TruSeq™ RNA and DNA sample preparation kits</i> (2010) (“Illumina TruSeq Data Sheet (2010)”)
1031	Craig <i>et al.</i> , <i>Identification of genetic variants using barcoded multiplexed sequencing</i> , NATURE METHODS 5:887-893 (2008) (“Craig (2008)”)
1032	Li & Durbin, <i>Fast and accurate short read alignment with Burrows–Wheeler transform</i> , BIOINFORMATICS 25:1754-1760 (2009) (“Li & Durbin (2009)”)
1033	Quail <i>et al.</i> , <i>Improved protocols for the Illumina Genome Analyzer sequencing system</i> , CURRENT PROTOCOLS IN HUMAN GENETICS 18.2.1-18.2.27 (2009) (“Quail (2009)”)

III. SUMMARY OF OPINIONS

3. In my opinion, Claims 1-16 of the '241 Patent are unpatentable as anticipated and/or obvious. My opinion is based on the following grounds:

- a. Ground 1: Claims 1-3, 5-7, 9-10, and 13-16 are anticipated under 35 U.S.C. § 102(a) by the '640 Patent Pub. (McCloskey).³
- b. Ground 2: Claims 1-3, 5-11, 13-16 are anticipated under 35 U.S.C. § 102(a) by the '202 Patent Pub. (Porreca).⁴
- c. Ground 3: Claims 1-3, 5-7, 9-12, and 14-15 are anticipated under 35 U.S.C. § 102(a) by the '389 PCT Appl. (Schmitt).⁵
- d. Ground 4: Claim 4 would have been obvious under 35 U.S.C. § 103 over the '202 Patent Pub. (Porreca) in view of Illumina (2008).⁶
- e. Ground 5: Claim 4 would have been obvious under 35 U.S.C. § 103 over the '389 PCT Appl. (Schmitt) in view of Illumina (2008).
- f. Ground 6: Claim 8 would have been obvious under 35 U.S.C. § 103 over the '202 Patent Pub. (Porreca) in view of Illumina (2008).

³ See Ex. 1005, McCloskey *et al.*, *Molecular encoding of nucleic acid templates for PCR and other forms of sequence analysis*, U.S. Patent Application Publication No US 2007/0020640 (2007) (the "'640 Patent Pub. (McCloskey)").

⁴ See Ex. 1006, Porreca *et al.*, *Methods and compositions for evaluating genetic markers*, U.S. Patent Application Publication No. US 2012/0165202 (the "'202 Patent Pub. (Porreca)").

⁵ See Ex. 1007, Schmitt *et al.* *Methods of lowering the error rate of massively parallel DNA sequencing using duplex consensus sequencing*, PCT International Publication No. WO 2013/142389 (2013) (the "'389 PCT Appl. (Schmitt)").

⁶ See Ex. 1008, Illumina sequencing, *Multiplexed sequencing with the Illumina Genome Analyzer system* (2008) ("Illumina (2008)").

- g. Ground 7: Claim 8 would have been obvious under 35 U.S.C. § 103 over the '389 PCT Appl. (Schmitt) in view of Illumina (2008).
- h. Ground 8: Claim 12 would have been obvious under 35 U.S.C. § 103 over the '202 Patent Pub. (Porreca) in combination with Li (2009).⁷

IV. EXPERIENCE AND QUALIFICATIONS

4. I am currently Founder, President, and Chief Executive Officer of RedVault Biosciences, LP (“RedVault”), which I founded in April 2013. RedVault is a biotechnology company committed to creating innovative technologies to advance genomic medicine. Before taking my position at RedVault, I was a tenured and adjunct Associate Professor in both the Department of Molecular & Human Genetics and at the Human Genome Sequencing Center (“HGSC”) at Baylor College of Medicine (“BCM”). I have also been an adjunct Associate Professor in the Cell & Molecular Biology Program at BCM and adjunct Assistant and Associate Professor in Chemistry at Rice University. At BCM, I worked as part of the team at the HGSC that contributed to the Human Genome Project.

⁷ See Ex. 1011, Li *et al.*, *The Sequence Alignment/Map format and SAMtools*, BIOINFORMATICS 25:2078-2079 (2009) (“Li (2009)”).

5. From September 2022 to September 2023, I was also Co-Founder and Chief Technical Officer of 454 Bio, Inc. 454 Bio is developing a portable DNA sequencer, a customizable device for decentralized, on-site sequencing.

6. In March 2002, I founded LaserGen, Inc. (“LaserGen”), which provides technology and reagents used in, among other things, nucleic acid sequencing methods. From March 2002 to April 2012, I was President and Chief Executive Officer of LaserGen, and from April 2012 to October 2012, I held the position of Chief Technology Officer. I was also a member of the Board of Directors from March 2002 to October 2012. From 2016 to 2018, Agilent Technologies acquired LaserGen for \$185M to convert its reversible terminator technology into a clinical diagnostic platform. I currently have no working relationship with LaserGen.

7. Since 1988, I have performed scientific research in the fields of molecular biology and chemical synthesis of nucleoside and nucleotide inhibitors. In particular, I have extensive experience with technology development in DNA sequencing methods including genomic DNA isolation, polymerase chain reaction (“PCR”), fragmenting genomic DNA, genomic DNA and complementary DNA (“cDNA”) library construction, bacterial cloning, nucleic acids chemistry, DNA modifying enzymes, polymerase engineering, fluorescent dyes, fluorescence imaging, and data analysis of multi-color images. I also have extensive experience

in the use and development of methods utilized to detect sequence variation, including single nucleotide variants (“SNVs”) and single nucleotide polymorphisms (“SNPs”), in all organisms, with a particular focus in humans and human immunodeficiency virus (“HIV”).

8. My research has been devoted in large part to developing next-generation sequencing (“NGS”) technologies, developing novel methods to study HIV transmission between individuals, and identifying molecular causes of novel forms of diabetes and their treatment. My colleagues and I have been deeply involved in PCR, DNA fragmentation, library construction, cloning, and the development of sequencing technologies. My industrial experience entails performing analytical chemical testing of inorganic materials (Aerojet-General), organic synthesis and characterization of porous polystyrene beads for applications in high performance liquid chromatography (Bio-Rad), development and applications of a robotic platform in Sanger sequencing (Applied Biosystems), and development and applications of high-throughput DNA sequencing to identify novel gene targets for small molecule screening (Merck Research Laboratories).

9. I received my Bachelor of Science degree in Biochemistry & Biophysics from the University of California, Davis in 1984. Between 1985 and 1988, I enrolled in a Master of Science program in the department of Chemistry at San Francisco State University. Although I completed the required coursework for

a Master of Science degree with an emphasis in Organic Chemistry, I elected not to complete this degree. Instead, in 1991, I enrolled in a doctoral program in the Department of Molecular and Human Genetics at Baylor College of Medicine in Houston, Texas. In 1996, I received my Ph.D. in Molecular and Human Genetics.

10. From 2000 to 2018, I gave lectures on sequencing technologies to graduate and medical students as part of the *Molecular Methods* course at BCM. These lectures were titled, *cDNA and Genomic Libraries*, *First-generation Sequencing*, and *Genotyping*, and *Next-generation Sequencing*. From 2001 to 2003, I also gave a lecture titled, *Mammalian Genome Analysis* in the *Mammalian Genomics* course at BCM.

11. I have also spoken at a variety of academic and industry conferences in the fields of molecular biology and DNA sequencing, including the *Next-Generation Sequencing Workshop* at Lübeck University in Germany; the *Centre de Regulació Genòmica* (“CRG”) Symposium, Barcelona Spain; the *Next-Generation Sequencing Conference* in Boston, MA; the *Advances in Genome Biology and Technology Meeting* (“AGBT”) in Marco Island, FL; the *Copenhagenomics Meeting* in Copenhagen, Denmark; the *American Society of Microbiology Conference* in San Francisco, CA; the SelectBio’s *Advances in Next Generation Sequencing Meeting*; and the *Critical Path to TB Drug Regimens* (“CPTR”) Workshop in Washington, DC.

12. I have authored 58 peer-reviewed papers and seven book chapters. Several of my review papers discuss emerging technologies and advances in DNA sequencing, including detection of sequence variations. I have written book chapters on topics related to polymerase chain reaction (“PCR”) and DNA sequencing. I have also reviewed numerous manuscripts for peer-review publications in the fields of molecular biology and DNA sequencing, including manuscripts applying the sequencing techniques described *supra*.

13. I have also conducted extensive research in the fields of molecular biology and DNA sequencing. My research has investigated and advanced fundamental techniques such as Sanger sequencing and NGS methods, chemical synthesis of nucleoside and nucleotide inhibitors, Sanger sequencing and NGS approaches for HIV forensics, and detection of genomic variation. Some of this research involved developing automated Sanger sequencing assays to measure mixed nucleotide ratios in DNA sequences of interest. In addition, my colleagues and I developed a novel diagnostic platform using intracellular and extracellular microRNAs for detection of various analytes. Like cell-free DNA, extracellular microRNAs are small nucleic acids that can be used as biomarkers for diagnostic purposes. My research has been funded by the National Institutes of Health (“NIH”), National Institute of Justice (“NIJ”), United States Department of Agriculture (“USDA”), and private foundations. Since 1999, I have received or

have been associated with numerous grants largely focusing on gene and genome sequencing totaling over \$170 million.

14. I am an inventor on 57 U.S. and European issued patents and patent applications directed towards molecular biology, DNA sequencing, and chemical synthesis. For example, I coinvented labelled and unlabeled cleavable terminating groups and methods for DNA sequencing described in U.S. Patent No. 8,148,503; 9,200,319 and 10,041,115. I also coinvented methods for the use of BODIPY fluorophore-labeled DNA for dye-primer sequencing described in U.S. Patent No. 5,861,287 and oligonucleotides labelled with BODIPY fluorophore compounds described in U.S. Patent No. 5,994,063. Further, I coinvented the pulsed-multiline excitation (“PME”) method in collaboration with the late Nobel Laureate Robert F. Curl, Ph.D., which resulted in U.S. Patent Nos. 6,995,841; 7,511,811; and 8,089,628. By way of collaboration with my scientific team at LaserGen, I also coinvented novel nucleotide terminators, which resulted in numerous patents including U.S. Patent Nos. 7,893,227; 7,897,737; 7,964,352; 8,148,503; 8,198,029; 8,361,727; 8,497,360; 8,877,905; 8,889,860; 8,969,535; 9,200,319; 9,399,798; 9,689,035; 10,041,115; and 11,001,886.

15. I have served on review panels for Genome Canada, Canadian Institute for Health Research (“CIHR”), National Aeronautics and Space Administration (“NASA”), and the U.S. Department of Energy (“DOE”), as well as various NIH

study sections for the National Cancer Institute (“NCI”), National Institute of Biomedical Imaging and Bioengineering (“NIBIB”), National Human Genome Research Institute (“NHGRI”), and National Institute for Allergy and Infectious Diseases (“NIAID”). In March 2018, I became a member of the College of Reviewers for CIHR.

16. From 2003 to 2006, I served on the editorial/advisory board for the journal *Genome Research*. From 2006 to 2012, I served as a scientific organizer for the AGBT Meeting. From 2011 to 2013, I served on the advisory committee of Genome Canada: *Advancing Technology Innovation Through Discovery* (“ATID”) projects – *Finding of Rare Disease Genes* in Canada (“FORGE Canada”), and the Canadian Pediatric Cancer Genome Consortium (“CPCGC”).

17. I have also served as an expert witness in several criminal trials in the United States involving the transmission of human immunodeficiency virus (HIV) between individuals. The criminal case involving Richard Schmidt was portrayed in an episode titled *Shot of Vengeance* on Forensic Files in 2003. The criminal case involving Philippe Padieu was reported on ABC News 20/20 by Elizabeth Vargas and on Oprah, both in 2009, and helped inspired an episode titled *Quickie* on Law & Order: Special Victims Unit in 2010.

18. I belong to the American Association for the Advancement of Science, the American Chemical Society, and the Texas Genetics Society. My professional

experience, education, and publications are presented in greater detail in my curriculum vitae, including a list of the cases in which I provided expert testimony within the last five years.⁸

19. As an expert in molecular biology, I am qualified to provide an opinion as to what one of ordinary skill in the art would have understood, known or concluded as of the 2013 time frame or several years prior to that.

V. TECHNICAL BACKGROUND

A. The relevant art

20. The '241 Patent is related to the field of “high throughput sequencing reactions” and discloses methods of detecting duplicate nucleic acid sequencing reads.⁹ The steps of sequencing a nucleic acid sample involves first preparing a library from the sample and then sequencing the library. A library preparation method typically has several processing steps that can also include amplifying sequences of interest using polymerase chain reaction (“PCR”). PCR generates multiple copies of (or “amplifies”) the nucleic acid sample so that there is a sufficient amount of material for the sequencing step. Because PCR amplification results in multiple copies (*i.e.*, duplicates) of the exact same sequence of nucleotides, it may be unclear whether those duplicates are PCR amplification

⁸ See Ex. 1016, Metzker CV.

⁹ See Ex. 1001, '241 Patent at 1:26-29.

products from the same DNA fragment or whether the same sequence independently exists from a different fragment from the same or from another portion of the genome.

21. It is desirable to identify which duplicates are merely PCR amplification products, as those copies do not provide useful information and will overrepresent the data. For instance, if a particular sequence contains a mutation, and that sequence is amplified ten-fold, then there will be ten copies of that mutation in the sequencing read data even though there was only a single occurrence of the mutation.

B. DNA

22. The information necessary to reproduce a given organism is found in its genome, which comprises the complete set of an organism's genes. The genome is comprised of strands of a polymer called *deoxyribonucleic acid* or "DNA." Each cell for a given organism typically contains one copy of the genome, called genomic DNA or "gDNA," which is found in the nucleus.

23. DNA strands are polymers of building blocks called *nucleotides*. A nucleotide is made up of three chemical components or groups: the nucleobase (commonly referred to as a base), the sugar, and one or more phosphate groups. The strands of DNA have *direction* and are commonly written in a 5'-to-3'

manner.¹⁰ For DNA, the specific sugar group is a 2'-deoxyribose. For ribonucleic acid ("RNA,") the specific sugar group is a ribose. In addition to this difference, DNA contains the base thymine ("T") while RNA contains the base uracil ("U"). Using an enzyme called a *polymerase*, nucleotides that are added during DNA synthesis are called 2'-deoxyribonucleoside-5'-triphosphates or "dNTPs," and those added during RNA synthesis are called ribonucleotide-5'-triphosphates or "NTPs" or sometimes "rNTPs."

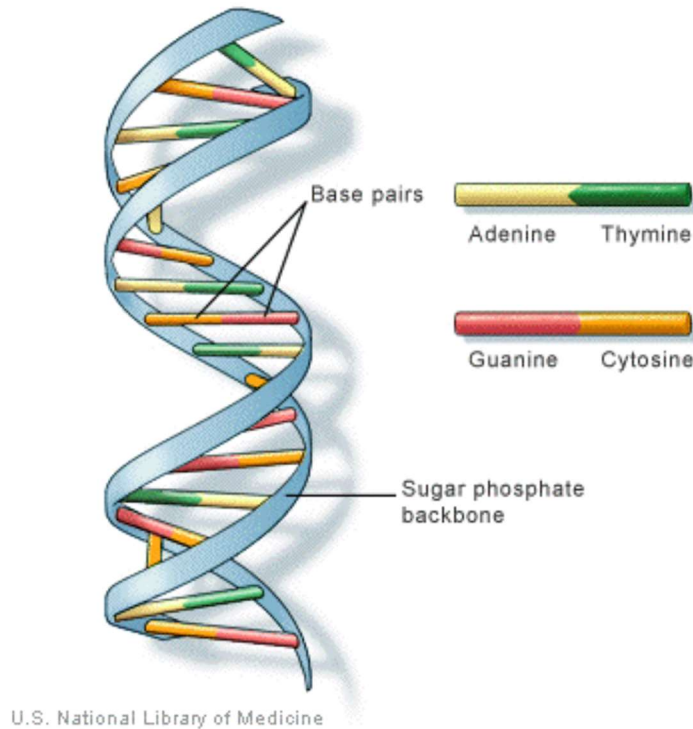
24. A polymerase that synthesizes DNA is called a DNA polymerase and one that synthesizes RNA is called a RNA polymerase. Polymerases can possess other activities in addition to synthesizing nucleic acids. For example, a polymerase can remove nucleotides from one or the other end of nucleic acids. Polymerases that remove nucleotides from the 5'-end of a nucleic acid has a 5'-exonuclease activity. Polymerases that remove nucleotides from the 3'-end of a nucleic acid has a 3'-exonuclease activity. Some polymerases can also add nucleotides to the ends of DNA fragments in a non-template manner, such as adding adenosine nucleotide ("A") to the end of a DNA fragment, called *A-tailing*.

25. The DNA structure can be characterized as a ladder. The two opposing legs of the DNA ladder are based on a repeated pattern of a sugar group and a

¹⁰ The numbering system of 5' and 3' are based on the position of carbon atoms on the sugar group.

phosphate group, called the DNA backbone. The chemical variance occurs in the rungs of the DNA ladder—*i.e.*, the four different bases: adenine (“A”), cytosine (“C”), guanine (“G”), and thymine (“T”)—which form hydrogen bonds according to specific base-pairing rules. “A” and “T” can form stable hydrogen bonds, and “C” and “G” can form stable hydrogen bonds. Said in another way, “A” does not pair with “G” and “C” does not pair with “T.” For RNA, “U” is substituted for “T” whereby “A” and “U” can form stable hydrogen bonds.

26. The formation of base-pairs between nucleotides on the two opposite strands causes the two strands to coil around each other to form a double helix structure, first described by James Watson and Francis Crick in 1953. Bases of nucleotides that follow these base-pairing rules are also called *Watson-Crick base-pairs* (“bp”). An example of a double helix is shown *infra*.



27. According to the base-pairing rules, stretches of nucleotides in one strand can form hydrogen bonds only with a complementary stretch of nucleotides in the other strand. Two single strands of DNA that come together by way of hydrogen bonding between stretches of complementary (or substantially complementary) nucleotides are said to be *hybridized* or *annealed* to one another. Because of the specificity of base-pairing, under most environmental conditions, a strand of DNA will have a specific stable complementary sequence. The hybridization of one nucleic acid strand (*i.e.*, DNA or RNA) with another nucleic acid strand (*i.e.*, DNA or RNA) has been widely exploited in the field of molecular biology and widely used to detect and identify nucleic acid sequences, as described *infra*.

28. This complementarity is critical for the heritability of genetic information. Because of the specificity of base-pairing, DNA polymerase can create or synthesize a complementary strand from an original single strand of a DNA called a *template* or *template molecule*. To start replication, DNA polymerase requires an *oligonucleotide* or *primer* (that is, a short DNA molecule typically between 15 and 30 bases) that will hybridize to a single-stranded template molecule. These replicated strands created by extending the primer with DNA polymerase are called *primer extension products*. As such, each strand of a double-stranded sequence can act as a template for the other. Cells are able to replicate DNA based on this property, and thus, pass exact copies of their genome to their progeny. This replication system has been widely exploited in the field of molecular biology to create one or more copies (also called *amplification*), such as the polymerase chain reaction (“PCR”).¹¹

C. PCR

29. A universal technique for replicating or amplifying nucleic acids in a laboratory setting is the polymerase chain reaction, or just “PCR.” This method

¹¹ See, e.g., Ex. 1018, Saiki *et al.*, *Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia*, SCIENCE 230:1350-1354 (1985) (“Saiki (1985)”); see also Ex. 1019, Saiki *et al.*, *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase*, SCIENCE 239:487-491 (1988) (“Saiki (1988)”).

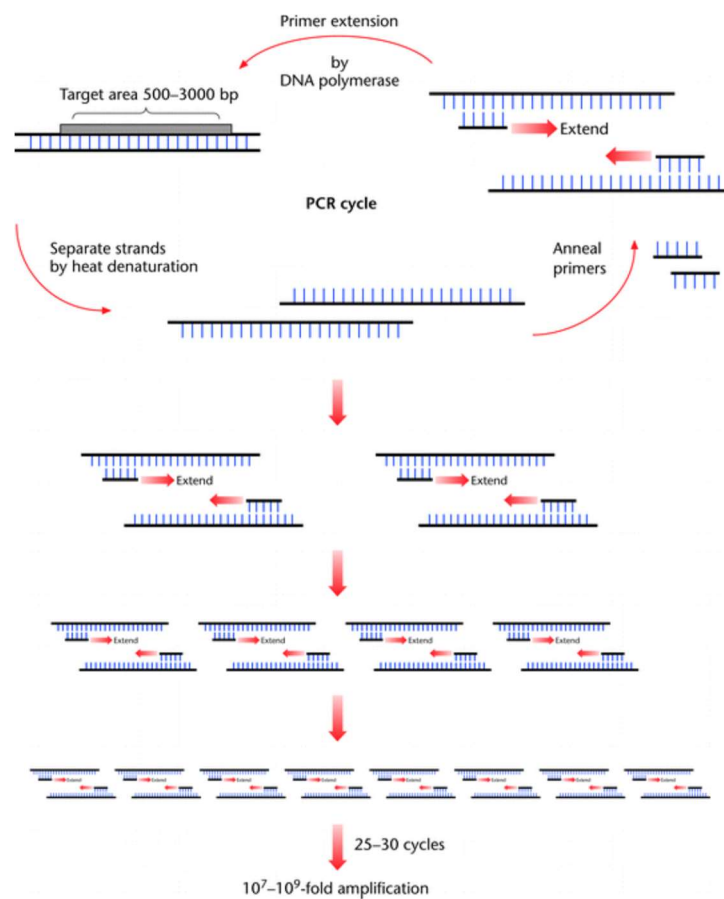
was invented by Dr. Kary Mullis in the 1980s.¹² PCR is used to specifically amplify a small amount of nucleic acid molecules, generating thousands to millions of copies of the target nucleic acid.¹³

30. Generally speaking, PCR involves repeated denaturation and replication of DNA sequences *in vitro* (i.e., outside of an organism). DNA polymerase copies or synthesizes the complementary strand from a single-stranded template. For this enzymatic reaction to occur, a partially double-stranded section of DNA is required. Typically, a primer hybridizes to a complementary region of a single-stranded template. Dr. Mullis and his co-inventors used short oligonucleotide primers complementary to the 3'-ends of the complementary strands of a sequence of interest. Because of the base-pairing rules, primer sequences can be designed to target specific nucleic acid sequences. DNA polymerase initiates synthesis of the nascent strand from the hybridized primers in a 5'-to-3' direction to create double-stranded DNA. Two primers are used for each

¹² See Ex. 1017, Metzker & Caskey, *Polymerase chain reaction*, In ENCYCLOPEDIA OF MEDICAL DEVICES AND INSTRUMENTATION, Second Edition, Volume 5, pp. 380-387 (2006) ("Metzker (2006)") at 380.

¹³ See generally Ex. 1018, Saiki (1985) and Ex. 1019, Saiki (1988).

template strand in PCR that define the targeted sequence of interest. The general concept of PCR is illustrated by the diagram *infra*.¹⁴



31. Going counterclockwise in the figure *supra*, the first step of a PCR reaction is to *denature* double-stranded DNA to yield single-stranded DNA. This typically is done by raising the reaction temperature to melt or by chemical means to separate double-stranded DNA or fragments into single strands. Following

¹⁴ See Ex. 1017, Metzker (2006) at Figure 1.

primer hybridization, the polymerase copies the template strand from the primer in a 5'-to-3' direction, making a complementary strand. Assuming optimal reaction conditions, the quantity of the original target sequence of interest, which is flanked by the two primers, will have been essentially doubled in one PCR cycle.¹⁵

32. Next, the PCR cycle is repeated multiple times. This is accomplished by denaturing the double-stranded template, annealing the primers, and extending the primers with DNA polymerase to synthesize new double-stranded templates. At each PCR cycle, the target sequence that lies between the primers (including the primer sequences themselves) will be replicated. Theoretically, a target sequence will be amplified exponentially as described by 2^n , where n is the number of cycles. In practice, however, the efficiency of the reaction varies, and decreases in later cycles. Regardless, PCR is able to create large quantities of DNA within a matter of hours.

33. PCR is a powerful tool for detecting human polymorphic variation that is associated with hereditary disease, aneuploidy, and cancer. Many techniques have been described to discriminate between wild-type and mutant alleles. For example, allele-specific PCR, also called competitive oligonucleotide priming¹⁶ or

¹⁵ *Id.* at 380. (“Each cycle doubles the region marked by the primer sequences.”).

¹⁶ *See* Ex. 1020, Gibbs *et al.*, *Detection of single DNA base differences by competitive oligonucleotide priming*, NUCLEIC ACIDS RESEARCH 17:2437-2448 (1989) (“Gibbs (1989)”).

called *amplification refractory mutation system*¹⁷ has been used for detecting polymorphic-specific markers. Multiplex PCR allows for the simultaneous amplification of multiple target regions and has been used to detect coding exon deletion(s) in X-linked disorders (these exons are gene sequences that are transcribed into messenger RNA (mRNA) and translated into one or more proteins). X-linked disorders include Duchenne muscular dystrophy¹⁸ and Lesch—Nyhan syndrome.¹⁹ Absence of specific PCR product(s) is diagnostic of exon deletion(s) in affected males, and “half-dosage” PCR products are diagnostic of carrier mothers. Other detection methods that involve scanning methods (*i.e.*, techniques that do not involve DNA sequencing) of PCR products for the presence of genetic alterations²⁰, genotyping methods, such as *mini-sequencing*, of PCR products to

¹⁷ See Ex. 1021, 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)”).

¹⁸ See Ex. 1022, Chamberlain *et al.*, *Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification*, NUCLEIC ACIDS RESEARCH 16:11141-11156 (1988) (“Chamberlain (1989)”).

¹⁹ See Ex. 1023, Gibbs *et al.*, *Multiplex DNA deletion detection and exon sequencing of the hypoxanthine phosphoribosyltransferase gene in Lesch-Nyhan families*, GENOMICS 7:235-244 (1990) (“Gibbs (1990)”).

²⁰ See Ex. 1024, Grompe, *The rapid detection of unknown mutations in nucleic acids*, NATURE GENETICS 5:111-117 (1993) (“Grompe (1993)”).

determine single nucleotide variations (“SNVs”),²¹ or as described *infra*, other sequencing technologies.

D. DNA sequencing technologies

34. DNA sequence technologies are important because they can help diagnose disease. The Human Genome Project begun in October 1990 as a joint effort of the NIH and DOE, with the ultimate goal of determining the complete sequence of the three billion nucleotides that make up the human genome.²² I was part of the project that was responsible for sequencing the first human genome at Baylor College of Medicine. It took approximately 13 years to complete the sequence of the first human genome. While the sequencing technologies I described below can produce entire genomes faster and cheaper than the first human genome, doctors or researchers that suspect certain DNA mutations may cause a particular disease will only focus on those part of the genome. Here, I first introduce sequencing technologies and then how those technologies are used to enrich one or more portions of the genome, called targets, to identify mutations that cause disease.

²¹ See Ex. 1025, Syvänen, *Accessing genetic variation: genotyping single nucleotide polymorphisms*, NATURE REVIEW GENETICS 2:930-942 (2011) (“Syvänen (2001)”).

²² I note this is the haploid state, and as diploid organism, the human genome is twice this size.

35. DNA sequencing includes several methods and technologies that are used to determine the order of nucleotide bases (that is, A, C, G, and T) in a DNA fragment. DNA sequencing methods can generally be characterized as either *first-generation sequencing*, which include Sanger sequencing,²³ or *next-generation sequencing* (“NGS”), discussed *infra*. In broad terms, next-generation sequencing technologies are distinguishable from first-generation sequencing technologies in that they produce substantially larger volumes of sequence information cheaply. I have reviewed many of the platforms using next-generation sequencing technologies.²⁴ I may rely on figures and information from this article.

36. There have been a number of NGS platforms that have been released commercially since 2005, including those sold by 454 Life Sciences Corporation—the GS20 and GS FLX® instruments,²⁵ Life Technologies—the SOLiD® instrument,²⁶ and Pacific Biosciences—the single-molecule real-time or SMRT® instrument.²⁷ Other NGS systems that have been commercialized include the Ion

²³ See Ex. 1026, Metzker, *Emerging technologies in DNA sequencing*, GENOME RESEARCH 15:1767-1776 (2005) (“Metzker (2005)”) at 1768-1770.

²⁴ See, e.g., Ex. 1010, Metzker, *Sequencing technologies — the next generation*, NATURE REVIEWS GENETICS 11:31-46 (2010) (“Metzker (2010)”).

²⁵ *Id.* at 39, Figures 1a, 3c, 3d.

²⁶ *Id.* at 36-39 at Figures 1a, 3a, 3b.

²⁷ *Id.* at 39 at Figures 1e, 4.

Torrent—the Personal Genome Machine (“PGM[®]”)²⁸ and Ion Proton[®] instruments and Oxford Nanopore—the MinION sequencer.²⁹ The Genome Analyzer (“GA”) I sequencer was developed by Solexa and released in late 2006. The GA II sequencer was developed after the acquisition of Solexa by Illumina in 2007. Since 2009, Illumina has released new instruments including multiple versions of HiSeq[®], MiSeq[®], NextSeq[®] and NovaSeq[®] systems. In this Declaration, I refer to a particular NGS technology as “platform-specific,” which refers to, at least, one of the instruments mentioned *supra*.

37. To perform NGS, a sequencing “library” is created from a nucleic acid sample. A library is the collection of DNA fragments that typically have been modified and amplified to be compatible with a given NGS sequencer.³⁰ To create genomic libraries, the DNA is typically sheared or enzymatically digested into smaller fragments, and platform-specific oligonucleotides, called adaptors³¹ are attached or ligated to both ends of the DNA fragment.³² These fragments can be

²⁸ See Ex. 1027, Rothberg *et al.*, *An integrated semiconductor device enabling non-optical genome sequencing*, NATURE 475:348-352 (2011) (“Rothberg (2011)”).

²⁹ See Ex. 1028, Jain *et al.*, *Improved data analysis for the MinION nanopore sequencer*, NATURE METHODS 12:351-356 (2015) (“Jain (2015)”).

³⁰ See Ex. 1010, Metzker (2010) at 32.

³¹ *Id.* In note the term “adaptor” is also commonly spelled as “adapter.” One of ordinary skilled in the art would have understood these spellings to mean this same thing.

³² See Ex. 1010, Metzker (2010) at 32.

amplified by a number of PCR approaches. Library preparation methods for NGS well-known and widely used by the time the first application leading to the '241 Patent.³³ By that time, in fact, standard kits to prepare sequencing libraries had been available for at least several years before the '241 Patent. These kits, like the asserted claims of the '241 Patent, employed platform-specific adaptors that contain sample-specific barcodes and sequences to which the sequencing primer hybridizes during the sequencing step. The kits were widely used with next-generation sequencers, which rendered use of such sequences a necessity for fast and cost-effective sequencing.

E. The use of molecular identifiers to detect PCR duplicates has been well-known in the art

38. Since at least the mid 2000s, and certainly before November 13, 2013, which for purposes of this Declaration I assume is the priority date of the '241 Patent,³⁴ the use of molecular identifiers to identify PCR duplicates was well understood in the art. For example, Miner (2004), discloses the use of “*molecular barcoding to label each genomic DNA template with an individual sequence tag*

³³ See, e.g., Ex. 1029, Illumina, *Multiplexing, Sample preparation guide* (2010) (“Illumina Manual (2010)”); see also Ex. 1030, Illumina, *TruSeq™ RNA and DNA sample preparation kits* (2010) (“Illumina TruSeq Data Sheet (2010)”); Ex. 1031, Craig *et al.*, *Identification of genetic variants using barcoded multiplexed sequencing*, NATURE METHODS 5:887-893 (2008) (“Craig (2008)”).

³⁴ For this IPR proceeding, I am not offering an opinion whether the '241 Patent is entitled to the priority date of November 13, 2013.

prior to PCR amplification.”³⁵ Miner (2004) recognized that there was “frequent observation of multiple amplified sequences derived from a single original molecule,” and taught the marking of each genomic fragment to allow for the identification of “redundant sequences.”³⁶ Miner (2004)’s method allowed for the “confirmation of the independent genomic origin of all sequences in final data sets derived from PCR amplification,”³⁷ and thus taught the identification of PCR duplicates as all sequences that derive from the same genomic origin. The other prior art references discussed *infra* also demonstrate that the ’241 Patent’s claimed method of identifying PCR duplicates was well-known in the art.

VI. APPLICABLE STANDARDS

39. I am not a lawyer, but I understand the following standards concerning the applicable law, as described *infra*.

³⁵ See Ex. 1004, Miner *et al.*, *Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR*, NUCLEIC ACIDS RESEARCH 32:e135 pp. 1-4 (2004) (“Miner (2008)”) at Abstract (emphasis added).

³⁶ *Id.* at 1.

³⁷ *Id.*

A. Relevant time period

40. As noted *supra*, I have assumed that the relevant time for my analysis of the pertinent art and level of ordinary skill in the art is on or before November 13, 2013, the earliest possible priority date for the '241 Patent.³⁸

B. Level of ordinary skill in the art

41. In my opinion, the relevant art is molecular biology. I have considered the level of ordinary skill based on my experience in the field, and the educational level of the inventor; the type of problems encountered in the art; prior art solutions to those problems; rapidity with which innovations are made; sophistication of the technology; and educational level of active workers in the field. I have been informed by counsel and understand that not all such factors may be present in every case, and one or more of these or other factors may predominate in this case. In the present case, I understand that the inventors are all Ph.D. level scientists that had 7-17 years of experience as of November 2013. Based on this and on my experience in working with others in the field, I believe that one of ordinary skill in the art in the relevant timeframe would have had (i) a Ph.D. in molecular biology, molecular genetics, bioinformatics, biochemistry, or a related field and at least two years of relevant applied research and/or industry experience in the field of nucleic acid

³⁸ For this IPR proceeding, I am not offering an opinion whether the '241 Patent is entitled to the priority date of November 13, 2013.

library preparation and sequencing experience, including but not limited to high-throughput sequencing technologies or (ii) a Master's degree in one of the same fields with at least four years of relevant applied research and/or industry experience in the field of nucleic acid library preparation and high-throughput sequencing technologies..

C. Anticipation

42. I have been informed by counsel and I understand that a patent claim is invalid as anticipated if a single piece of prior art discloses each and every limitation of that claim.

43. I have been informed by counsel and understand that a single piece of prior art may anticipate without disclosing a particular limitation if that feature is necessarily present, *i.e.*, inherent, in that piece of art. When an anticipating reference is silent about an asserted inherent characteristic, such gap in the reference may be filled with recourse to extrinsic evidence. Such evidence must make clear that the missing descriptive matter is necessarily present in the thing described in the reference, and that it would be so recognized by one of ordinary skill in the art. I further understand that a failure of the skilled artisan to contemporaneously recognize an inherent property, function or ingredient of a prior art reference does not preclude a finding of anticipation.

D. Obviousness

44. I have been informed by counsel and understand that an invention cannot be patented if the subject matter as a whole would have been obvious at the time the invention was made to one of ordinary skill in the art to which the subject matter pertains. I understand that something can be obvious over a single reference in combination with the knowledge of one of ordinary skill in the art, or can come from the combination of sources. I also understand that the combination of familiar elements according to known methods that yield predictable results are likely obvious. I have also been informed by counsel and understand that in an obviousness analysis it is assumed that the skilled artisan had knowledge of all prior art in the field of endeavor relevant to the claim. Where prior art falls outside that field, it should still be considered if it is reasonably pertinent to the particular problem with which the inventor was involved.

45. I have been informed by counsel and understand that the obviousness determination turns on underlying factual inquiries involving: *(i)* the scope and content of the prior art; *(ii)* differences between the claims and the prior art; *(iii)* the level of ordinary skill in the pertinent art; and *(iv)* secondary considerations of non-obviousness. Secondary considerations of non-obviousness include, among other things, commercial success, long felt but unsolved needs, and failure of others. I have been informed by counsel and understand that a party seeking to invalidate a

patent as obvious must demonstrate by clear and convincing evidence that a skilled artisan would have been motivated to combine the teachings of the prior art references to achieve the claimed invention, and that one of ordinary skill in the art would have had a reasonable expectation of success in doing so. I further understand that a patent may be invalid as obvious to try if the options that one of ordinary skill would have encountered were finite, small, or easily traversed, and one of ordinary skill in the art would have had a reason to select the route that produced the claimed invention. I understand that this motivation from the art can be explicit or implicit using a skilled artisan's common sense.

46. In addition, I have been informed by counsel and understand that obviousness requires more than a showing that the prior art includes separate references covering each separate limitation in a claim. Rather, obviousness requires the additional showing that one of ordinary skill in the art at the time of the invention would have selected and combined those prior art elements in the normal course of research and development to yield the claimed invention. I also understand that it is impermissible to evaluate obviousness from a hindsight perspective, using the teachings of the patent(s) at issue as a guide. One of ordinary skill in the art is not an automaton, however, and is assumed that the skilled artisan would have had a reasonable degree of creativity, and an analysis of obviousness may take account of the inferences and creative steps that the skilled artisan would

employ. The analysis of obviousness can take into account the interrelated teachings of several patents or sources, incentives to change the designs in the prior art, market forces driving such changes, background knowledge, and the problems known in the art at the time. A combination of old elements that perform the same functions that they had been known to perform that yields no more than one would expect from such an arrangement is obvious.

VII. THE '241 PATENT

47. The '241 Patent is titled, *Compositions and methods for identification of a duplicate sequencing read*. The '241 Patent issued from Application No. 17/381,888 (the "'888 Application"), which was filed on July 21, 2021.³⁹ The '241 Patent claims priority to U.S. Provisional Application No. 61/903,826 (the "'826 Provisional"), which was filed on November 13, 2013.

48. The '241 Patent "relates generally to the field of high throughput sequencing reactions and the ability to discern artifacts arising through sequence duplications from nucleotide molecules that are unique molecules."⁴⁰ The '241 Patent states that "[i]n RNA sequencing applications, accurate gene expression measurements may be hampered by PCR duplicate artifacts that occur during

³⁹ See Ex. 1001, '241 Patent at cover page.

⁴⁰ *Id.* at 1:26-29.

library amplification.”⁴¹ When analyzing sequencing data, “when two or more identical sequences are found, it can be difficult to know if these represent unique cDNA molecules derived independently from different RNA molecules, or if they are PCR duplicates derived from a single RNA molecule.”⁴² The ’241 Patent further states that in some applications such as genotyping by sequencing, “duplicate reads can be considered non-informative and may be collapsed down to a single read, thus reducing the number of sequencing reads used in final analysis.”⁴³

49. The ’241 Patent purports to provide “novel methods for identifying true duplicate reads during sequencing, such as to improve data analysis of sequencing data, and other related advantages.”⁴⁴ To identify duplicates, the ’241 Patent states the use of an “identifier site” that can be included in an “adaptor” appended to the target nucleotide to be sequenced.⁴⁵ The ’241 Patent states that an adaptor is “an oligonucleotide sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target

⁴¹ *Id.* at 1:34-36.

⁴² *Id.* at 1:36-41.

⁴³ *Id.* at 1:41-45.

⁴⁴ *Id.* at 2:6-9.

⁴⁵ *Id.* at 4:16-34.

polynucleotide strand of interest.”⁴⁶ An identifier site “is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads.”⁴⁷ An example of the application of an adaptor containing an identifier site is shown in the ’241 Patent’s Figure 2A, as illustrated *infra*.

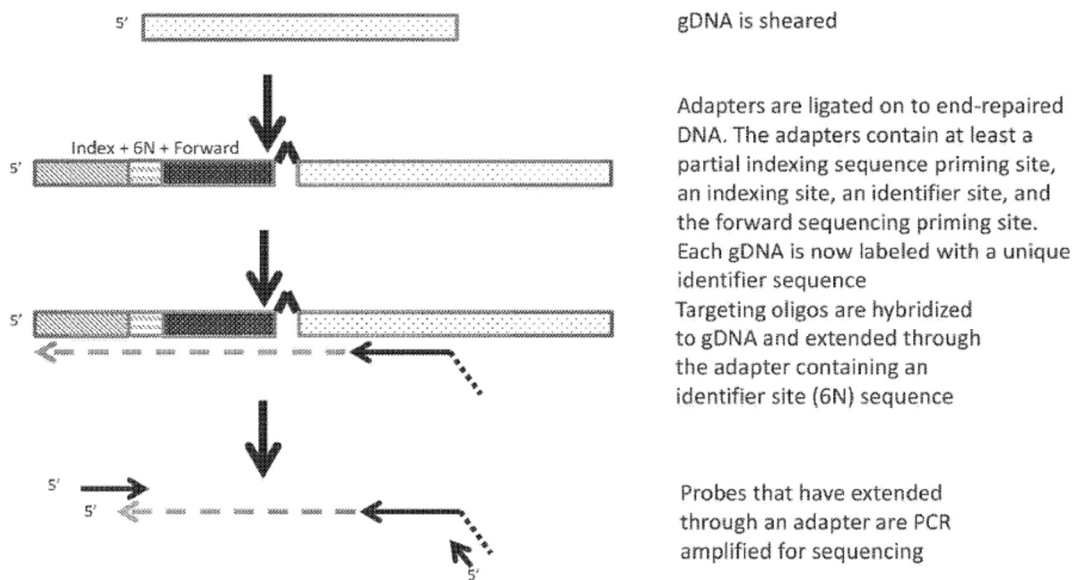


Figure 2A

50. As shown *supra*, Figure 2A “[t]he adaptor contain at least a partial indexing sequence priming site, an indexing site, an identifier site, and the forward sequencing priming site” that is “ligated on to end-repaired DNA.”⁴⁸ The ’241 Patent further states that “a sequencing read is identified as a duplicate

⁴⁶ *Id.* at 4:17-21.

⁴⁷ *Id.* at 4:45-48.

⁴⁸ *Id.* at Figure 2A.

sequencing read when it contains an identifier site and target sequence that is the same as another sequencing read from the same population of sequencing reads.”⁴⁹ The ’241 Patent states that, once duplicate reads are identified, they can be removed using common tools such as SAMtools. “Duplicate reads can be filtered using ‘samtools rmdup’, wherein reads with identical external coordinates are removed, only one read with highest mapping quality is retained. After filtering, a filtered set of deduplicated reads can be used in any downstream analysis.”⁵⁰

51. The claims of the ’241 Patent recite methods for identifying PCR duplicates. Claim 1 of the ’241 Patent is reproduced *infra*:

A method for detecting duplicate sequencing reads, the method comprising:
obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor, wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment;
sequencing the amplicons to generate sequence reads that include identifier and target sequences; and
identifying sequence reads with identical identifier and target sequences as duplicates.⁵¹

⁴⁹ *Id.* at 14:2-6.

⁵⁰ *Id.* at 14:49-53.

⁵¹ *Id.* at 25:33-43.

VIII. CLAIM CONSTRUCTION

52. I understand that Petitioner has taken the position that the term “adaptor” means “any oligonucleotide that is attached to the target sequence.” I have reviewed the specification of the ’241 Patent and understand that the its specification defines “adaptor” as “an oligonucleotide sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest.”⁵² I am also aware that during prosecution of the ’357 Patent,⁵³ the parent of the ’241 Patent that shares the same specification, the examiner reviewed the above definition of “adaptor” and concluded that “since any oligonucleotide can perform such function, any oligonucleotide is considered to anticipate this term.”⁵⁴

53. I further understand that the specification states that “appending the adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction.”⁵⁵ I also understand that Claim 10 of the ’241 Patent recites “wherein the adaptors have been appended to

⁵² *Id.* at 4:20-25.

⁵³ See Ex. 1002, Amorese *et al.*, *Compositions and methods for identification of a duplicate sequencing read*, U.S. Patent No. 11,098,357 (2020) (the “’357 Patent”).

⁵⁴ See Ex. 1013, ’357 Patent File History at 0077.

⁵⁵ See Ex. 1001, ’241 Patent at 9:34-37.

the fragments using a ligation reaction or a priming reaction.”⁵⁶ An adaptor according to the ’241 Patent thus can be appended *via* a ligation reaction or a priming reaction, and is not limited to only those that are ligated to the target sequence. Taking all of this into account, I agree with Petitioner that, in the context of the ’241 Patent, an “adaptor” means “any oligonucleotide that is attached to the target sequence.” I have applied this construction of “adaptor” in my analysis *infra*. For the remaining claim terms, I have applied their plain and ordinary meaning.

IX. PRIOR ART

54. I have reviewed the prior art references listed *infra*.

A. The ’640 Patent Pub. (McCloskey)

55. The ’640 Patent Pub. (McCloskey) is a U.S. Patent Application Publication titled, *Molecular encoding of nucleic acid templates for PCR and other forms of sequence analysis*.⁵⁷ It was filed on July 21, 2005 and published on January 25, 2007.⁵⁸ The ’640 Patent Pub. (McCloskey) is directed to “methods of authenticating a nucleic acid molecule and its sequence with a molecular barcode and batch-stamp,”⁵⁹ which allows for the identification of duplicates as well as the sample from which the DNA originated.

⁵⁶ *Id.* at Claim 10.

⁵⁷ *See* Ex. 1005, ’640 Patent Pub. (McCloskey) at cover page.

⁵⁸ *Id.*

⁵⁹ *Id.* at Abstract.

56. The '640 Patent Pub. (McCloskey) explains that PCR can generate multiple copies of DNA sequences from a limited amount of DNA, but that such reactions can “result in a *skewed yield of PCR products such that there is a high degree of redundancy* for a small portion of the original genomic sequences.”⁶⁰ To address this problem, the '640 Patent Pub. (McCloskey) discloses the tagging of individual DNA molecules with a “sequence tags,” which include barcodes to distinguish redundant sequences.⁶¹ One of ordinary skill in the art would have understood that redundant sequences are duplicate sequences. Specifically, the '640 Patent Pub. (McCloskey) teaches an adaptor containing both a “random barcode” and a “batch-stamp” that is attached to a target DNA sequence prior to PCR amplification.⁶² The barcoding allows for the identification of redundant sequences,⁶³ and the batch-stamp identifies which sample each sequence originated

⁶⁰ *Id.* at [0003] (emphasis added).

⁶¹ *Id.* at [0106]. “Here, a similar concept was applied to *the labeling of individual genomic fragments with distinct sequence tags. The ability to barcode and “batchstamp” genomic DNA sequences from individual alleles is useful in situations where the amount of template DNA is limited, thus identifying contaminants and redundant sequences* arising from template re-cloning.” (emphasis added).

⁶² *Id.* at [0060].

⁶³ *Id.* at [0005]; *see also id.* at [0016]. *Redundant hypermethylated sequences recovered from independent bacterial colonies (SEQ ID NO:4 and SEQ ID NO:5), with identical barcodes and methylation patterns*; D: A hypomethylated sequence with a distinctive barcode (SEQ ID NO:6); E, F: *Redundant hypomethylated sequences with identical barcodes recovered from independent*

from.⁶⁴ The '640 Patent Pub. (McCloskey) further teaches the removal of duplicate reads.⁶⁵

57. The '640 Patent Pub. (McCloskey) discloses two types of embodiments using barcoded adapters. The first type is shown in Figure 1 of the '640 Patent Pub. (McCloskey), as illustrated *infra*.

bacterial colonies (SEQ ID NO:7 and SEQ ID NO:8). These are distinguishable as redundant and as different from the hypomethylated sequence D only because of barcoding; . . . Redundant hypermethylated sequences are denoted with asterisks (*), and redundant hypomethylated sequences with plus signs (+).") (emphases added).

⁶⁴ *Id.* at [0016]; *see also id.* at [0027]. ("The term "experimental identification information" or "***batch-stamp***" refers to information that uniquely identifies a specific experiment. For example, a batch-stamp may identify a sample, a patient, and/or an analysis date.") (emphasis added).

⁶⁵ *Id.* at [0063]. ("***A redundant sequence was also recovered, meaning that a MLM1-positive sequence with an already seen bar-code was recovered. One of these two sequences was counted as a valid sequence.***") (emphasis added).

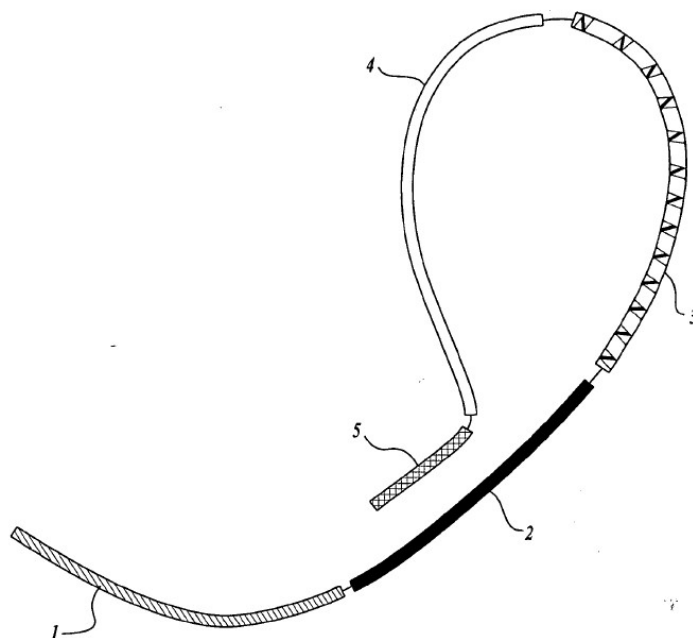


Fig.1.

Figure 1 of the '640 Patent Pub. (McCloskey) shows an adaptor containing several sequences labeled by numbers 1-5. These sequences are: (1) “a sequence complementary to the target nuclei[c] acid molecules, (2) “a batch-stamp sequence,” (3) “a random barcode,” (4) “a leftward primer binding site,” and (5) a 5' tethering sequence.”⁶⁶ Sequence (1) hybridizes to the target nucleic acid strand. DNA polymerase then extends sequence (1) along the target strand,⁶⁷ generating a complete strand that is complementary to the target sequence and that is attached to

⁶⁶ *Id.* at [0013].

⁶⁷ *Id.*, e.g., at [0058]. (“***Oligo MLM1 was denatured, annealed to genomic DNA, and extended*** as described in EXAMPLE 1.”) (emphasis added).

an adaptor containing the sequences numbered 1-5, described *supra*. The resulting product is then PCR amplified, as shown in Figures 2A-D, and sequenced.⁶⁸ The '640 Patent Pub. (McCloskey) further teaches that redundant sequences identified by this method are removed. "A redundant sequence was also recovered, meaning that a MLM1-positive sequence with an already seen bar-code was recovered. ***One of these two sequences was counted as a valid sequence.***"⁶⁹ One of ordinary skill in the art would have understood that counting one of two sequences would require discarding or removing the other sequence. The '640 Patent Pub. (McCloskey) teaches several examples that utilize a hairpin linker containing both the barcode and batch-stamp.⁷⁰ One of ordinary skill in the art would have understood that a hairpin linker is commonly used as an adapter sequence.

58. The '640 Patent Pub. (McCloskey)'s second type of embodiment that utilizes a hairpin adapter is shown in Figure 3, as illustrated *infra*

⁶⁸ *Id.*, e.g., at [0056]-[0061].

⁶⁹ *Id.*, e.g., at [0063].

⁷⁰ *Id.* at [0092]-[0106] (Example 6).

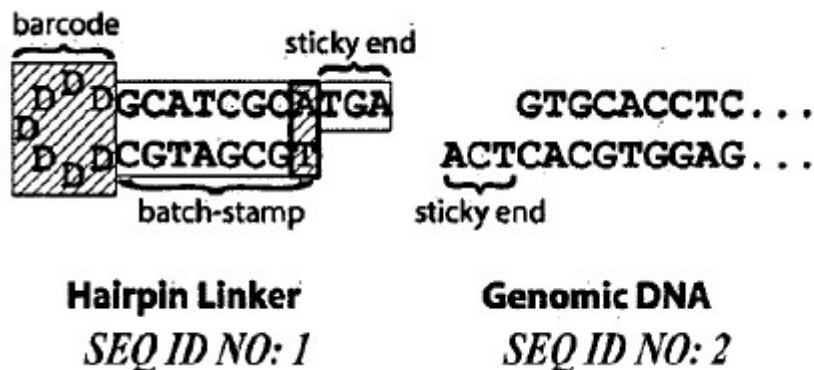


Fig.3.

Example 6 in the '640 Patent Pub. (McCloskey) describes the use of the hairpin adaptor to identify redundant or duplicate sequences.⁷¹ Example 6 in the '640 Patent Pub. (McCloskey) involves the ligation of a hairpin linker adaptor containing a barcode and a batch-stamp onto a fragment of genomic DNA.⁷² The '640 Patent Pub. (McCloskey) explains that “[i]n response to the challenges of PCR *redundancy* and contamination associated with PCR amplification of limited amounts of DNA template, *genomic DNA fragments were labeled with molecular sequence barcodes and ‘batchstamps’ prior to PCR amplification by including these molecular labels in the hairpin linker sequence (FIG. 3)* that is used in hairpin-bisulfite PCR.”⁷³ The resulting product is amplified by PCR and then sequenced.⁷⁴ The sequencing results include reads of the barcode, batch-stamp, and

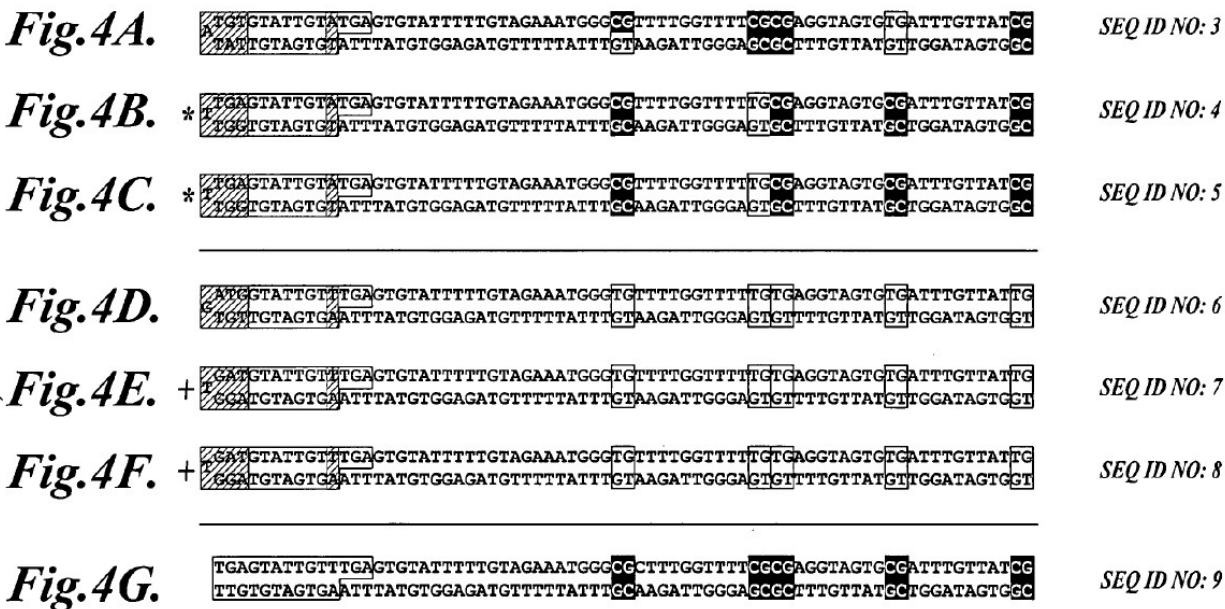
⁷¹ *Id.*

⁷² *Id.* at [0095].

⁷³ *Id.* at [0093] (emphases added).

⁷⁴ *Id.* at [0098], [0103].

target sequences, and are shown *infra* in Figures 4A-G of the '640 Patent Pub. (McCloskey).⁷⁵



B. The '202 Patent Pub. (Porreca)

59. The '202 Patent Pub. (Porreca) is a U.S. Patent Application Publication, titled *Methods and compositions for evaluating genetic markers*.⁷⁶ The '202 Patent Pub. (Porreca) was filed on March 13, 2012 and published on June 28, 2012.⁷⁷ The '202 Patent Pub. (Porreca) relates to next generation sequencing (NGS) techniques, specifically, the sequencing and analysis of nucleic acids and the disclosure of various methods of preparing nucleic acid libraries and sequencing

⁷⁵ See also the discussion at [0103] and [0106] in the '640 Patent Pub. (McCloskey).

⁷⁶ See Ex. 1006, '202 Patent Pub. (Porreca) at cover page.

⁷⁷ *Id.*

regions of interest. The '202 Patent Pub. (Porreca) discloses “methods and compositions that are useful to reduce bias and increase the reproducibility of multiplex analysis of genetic loci.”⁷⁸

60. The '202 Patent Pub. (Porreca) explains that certain biases can arise in sequencing analysis, including biases from PCR amplification. “[C]ertain types of genetic information can be under-represented or over-represented in a genetic analysis due to *a combination of stochastic variation and systematic bias* in any of the preparative stages (e.g., capture, *amplification*, etc.), determining stages (e.g., allele-specific detection, *sequencing*, etc.), data interpretation stages (e.g., determining whether the assay information is sufficient to identify a subject as homozygous or heterozygous), and/or other stages.”⁷⁹ To address biases associated with PCR amplification, the '202 Patent Pub. (Porreca) teaches the use of differentiator tag sequences, that “uniquely tag individual nucleic acid molecules.”⁸⁰ The '202 Patent Pub. (Porreca) teaches that “sequence tags (referred to as differentiator tags)” are “*useful to determine whether target nucleic acid sequences identified in an assay are* from independently isolated target nucleic acids or *from multiple copies of the same target nucleic acid molecule (e.g., due*

⁷⁸ *Id.* at Abstract.

⁷⁹ *Id.* at [0008] (emphasis added).

⁸⁰ *Id.* at [0023].

*to bias in a preparative step, for example, amplification).*⁸¹ One of ordinary skill in the art would have understood that multiple copies of sequence reads from the same target nucleic acid molecules are duplicate sequence reads.

61. The '202 Patent Pub. (Porreca) teaches several methods of incorporating the differentiator tag.⁸² For example, “nucleic acids comprising differentiator tag sequences may be incorporated by ligation.”⁸³ “Alternatively, *the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases)*, so that only one ligation must be performed.”⁸⁴ The differentiator tag may also be incorporated into a PCR primer “such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order).”⁸⁵ Figure 4C of the '202 Patent Pub. (Porreca) shows an example of an adaptor containing a differentiator tag sequence (NNNN) that is appended to the target nucleic acid molecule of interest, as illustrated *infra*.

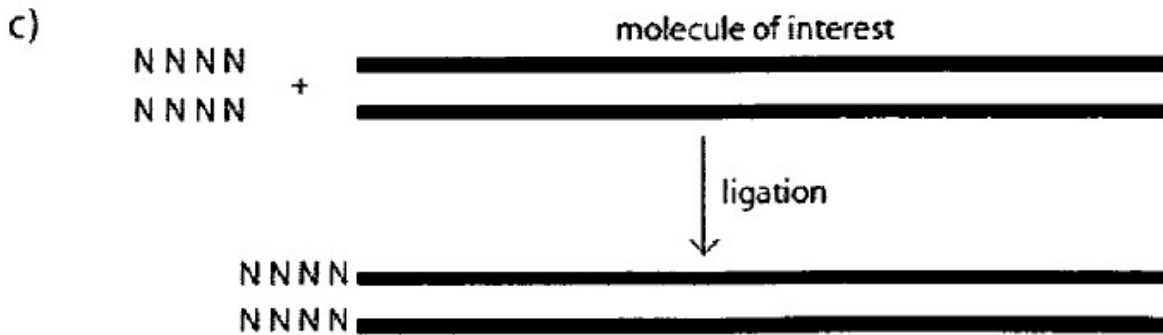
⁸¹ *Id.* at [0024] (emphases added).

⁸² *Id.* at [0150].

⁸³ *Id.*

⁸⁴ *Id.* (emphasis added).

⁸⁵ *Id.*



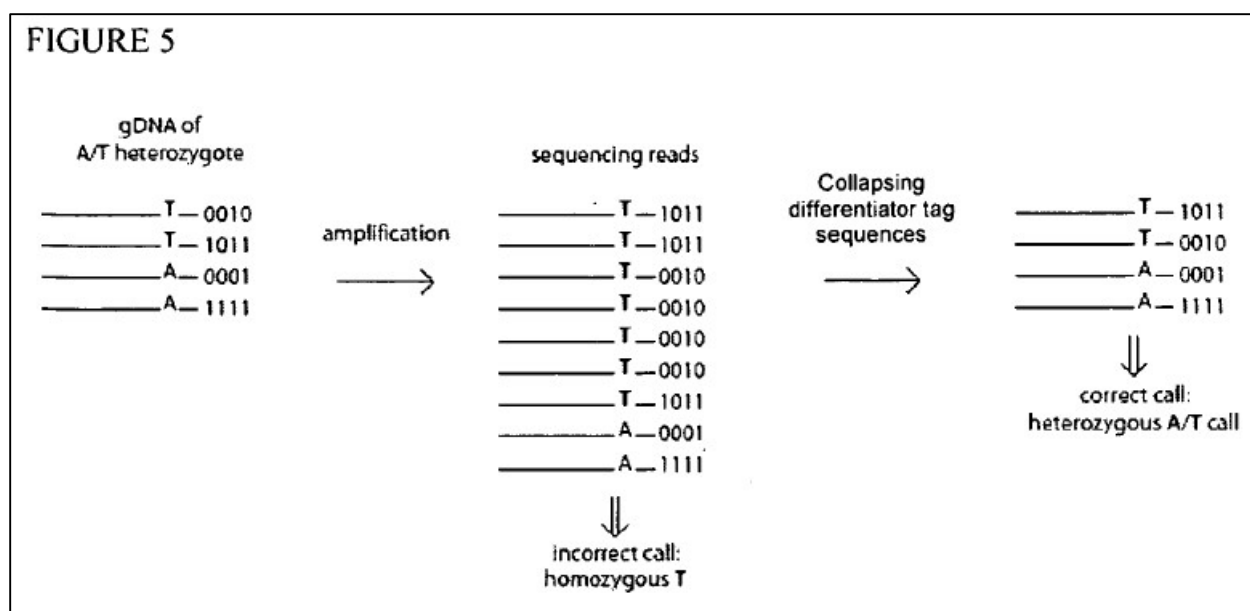
The '202 Patent Pub. (Porreca) also explains that “***FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.***”⁸⁶

62. The '202 Patent Pub. (Porreca) further teaches that sequence reads having the same target sequence and differentiator tag sequence are identified and collapsed. Specifically, the '202 Patent Pub. (Porreca) explains that “***[f]or any target:differentiator tag combination observed more than once, all such reads are ‘collapsed’ into a single read.***”⁸⁷ Figure 5 of the '202 Patent Pub. (Porreca) depicts this process, as illustrated *infra*.

⁸⁶ *Id.* at [0140] (emphases added).

⁸⁷ *Id.* at [0298] (emphasis added).

FIGURE 5



63. Figure 5 of the '202 Patent Pub. (Porreca) shows several PCR duplicates resulting from the amplification step. The sequencing reads in the middle portion of the figure show that there are three copies of sequence reads associated the "1011" differentiator tag and four copies of sequence reads associated the "0010" tag. These duplicate sequence reads are removed from the final read after the step "[c]ollapsing differentiator tag sequences," which results in a single read for each differentiator tag sequence. One of ordinary skill in the art would have understood that collapsing identical differentiator tag sequences in a single read would require discarding or removing the other sequences in that group. I note that the '241 Patent also acknowledges that collapsing identical reads into a single read results in removing duplicate sequence reads. "[D]uplicate reads can be considered non-informative and *may be collapsed down to a single read, thus reducing the*

number of sequencing reads used in final analysis.”⁸⁸ The ’202 Patent Pub. (Porreca) further teaches the use of “barcodes” that can “uniquely identify individual samples (e.g., patient samples) that may be mixed together.”⁸⁹

C. ’389 PCT Appl. (Schmitt)

64. The ’389 PCT Appl. (Schmitt) is a PCT application titled “*Methods of lowering the error rate of massively parallel DNA sequencing using duplex consensus sequencing.*”⁹⁰ It was filed on March 15, 2013 and was published on September 26, 2013.⁹¹ The ’389 PCT Appl. (Schmitt) is also directed to NGS techniques and teaches methods of addressing errors and bias associated with NGS, including through the use of small molecule identifiers (“SMIs”).

65. Like the references discussed *supra*, the ’389 PCT Appl. (Schmitt) teaches a method of identifying PCR duplicates, here through the use of a “single molecule identifier (SMI) adaptor molecule” that can be attached to the DNA

⁸⁸ See Ex. 1001, ’241 Patent at 1:42-45 (emphasis added).

⁸⁹ See Ex. 1006, ’202 Patent Pub. (Porreca) at [0149]. (“In addition, ***barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together.***”) [emphasis added, citing to the ’640 Patent Pub. (McCloskey)]. See also *id.* at [0102]. (“In some embodiments, a plurality of captured nucleic acids from each one of a plurality of patient samples may be combined in a single multiplex analysis reaction. In some embodiments, ***samples from different subjects are tagged with subject-specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject.***”) (emphasis added).

⁹⁰ See Ex. 1007, ’389 PCT Appl. (Schmitt) at cover page.

⁹¹ *Id.*

fragments being sequenced and that uniquely identifies each fragment.⁹² This SMI adaptor molecule is shown in Figure 1 of the '389 PCT Appl. (Schmitt), as illustrated *infra*.

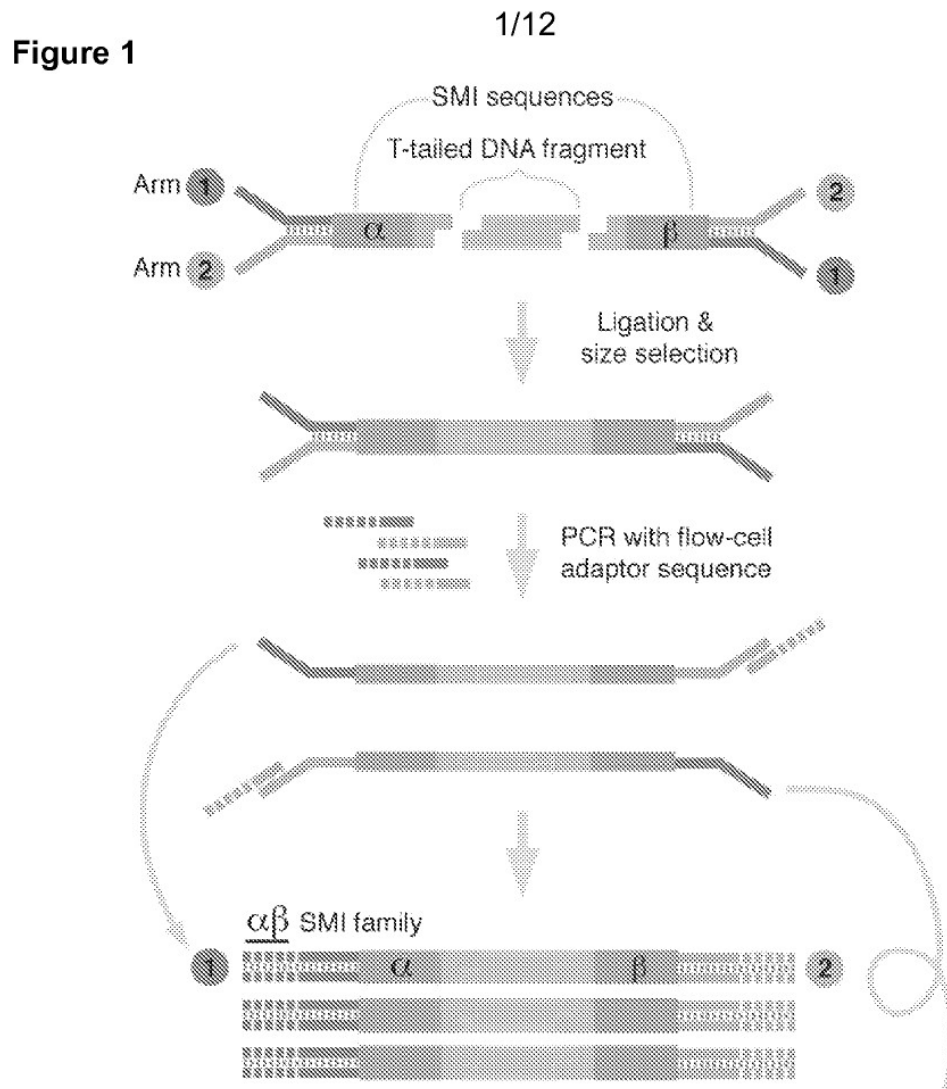


Figure 1 of the '389 PCT Appl. (Schmitt) shows that an adaptor containing an SMI sequence (represented as α and β segments) is ligated to the target sequence. The

⁹² *Id.* at [0009].

'389 PCT Appl. (Schmitt) explains that the resulting product then undergoes "PCR amplification with primers containing Illumina flow-cell-compatible tails" to "generate families of PCR duplicates."⁹³ The resulting amplified products are then sequenced, as illustrated *supra* in the bottom portion of Figure 1 of the '389 PCT Appl. (Schmitt).

66. The '389 PCT Appl. (Schmitt) further teaches that "[b]ecause the use of SMI tags (or 'double-stranded SMI sequences') allows every molecule to be uniquely labeled prior to PCR duplication, *true PCR duplicates may be unambiguously identified by virtue of having a common (i.e., the same or identical) SMI sequence.*"⁹⁴ Because it is desirable to "restrict sequencing analysis to uniquely sequenced DNA fragments, *many sequencing studies include a step to filter out PCR duplicates.*"⁹⁵ The '389 PCT Appl. (Schmitt) eliminates PCR duplicates by grouping reads with common target and SMI sequences and collapsing them to "generate a consensus read."⁹⁶ As described *supra*, one of ordinary skill in the art would have understood that collapsing identical sequences

⁹³ *Id.* at [0012].

⁹⁴ *Id.* at [0085] (emphasis added).

⁹⁵ *Id.* (emphasis added).

⁹⁶ *Id.* at [0096]. ("*Reads having common (i.e., identical) SMI sequences were grouped together, and were collapsed to generate a consensus read.*") (emphasis added).

having common target and SMI sequences in a single consensus read would require discarding or removing the other sequences in that group. I note that the '241 Patent also acknowledges that collapsing identical reads into a single read results in removing duplicate sequence reads. “[D]uplicate reads can be considered non-informative and *may be collapsed down to a single read, thus reducing the number of sequencing reads used in final analysis.*”⁹⁷

D. Li (2009)

67. Li (2009) is titled, *The Sequence Alignment/Map format and SAMtools* and was published on June 8, 2009.⁹⁸ I have been informed by counsel and understand that Li (2009) is prior art to the '241 Patent under 35 U.S.C. 102(a).

68. Li (2009) describes the “Sequence Alignment/Map (SAM)” format, which is “a generic alignment format for storing read alignments against reference sequences” and supports short and long reads “produced by different sequencing platforms.”⁹⁹ Li (2009) explains that different novel sequencing technologies have produced a “variety of new alignment tools” that generate alignments in different formats, which can complicate downstream processing.¹⁰⁰ Li (2009) further describes a need in the art for a “common alignment format that supports all

⁹⁷ See Ex. 1001, '241 Patent at 1:42-45 (emphasis added).

⁹⁸ See Ex. 1011, Li (2009) at 2078.

⁹⁹ *Id.* at Abstract.

¹⁰⁰ *Id.* at 2078.

sequence types and aligners” and “a well-defined interface between alignment and downstream analyses, including variant detection, genotyping and assembly.”¹⁰¹ Li (2009) discloses SAM as a format that achieves these goals.

69. Li (2009) also introduces SAMtools, which is “a library and software package for parsing and manipulating alignments in the SAM/BAM format.”¹⁰² SAMtools can perform various tasks, including “convert from other alignment formats, sort and merge alignments, *remove PCR duplicates*,” and “show alignments in a text-based viewer.”¹⁰³ As Li (2009) involves sequence reads generated by a variety of NGS platforms, one of ordinary skill in the art would have understood the phrase *remove PCR duplicates* to mean removing duplicate sequence reads from amplicons created by PCR. Li (2009) accomplishes this task by identifying “*the position of the first aligned base*” of the sequence read mapped to a reference genome.¹⁰⁴ By November 13, 2013, the use of SAMtools was widely used in the NGS community for removing duplicate sequence reads from amplicons created by PCR, which is acknowledged in the ’241 Patent.¹⁰⁵

¹⁰¹ *Id.*

¹⁰² *Id.* at 2079.

¹⁰³ *Id.* at Figure 1 legend (emphasis added).

¹⁰⁴ *Id.* (emphasis added).

¹⁰⁵ See Ex. 1001, ’241 Patent at 14:49-53. (“***Duplicate reads can be filtered using ‘samtools rmdup’***”, wherein reads with identical external coordinates are

E. *Illumina (2008)*

70. Illumina (2008) is a datasheet published by *Illumina titled, Multiplexed sequencing with the Illumina Genome Analyzer system*.¹⁰⁶ The Illumina (2008) datasheet is dated December 2, 2008.¹⁰⁷ This reference explains Illumina's multiplex sequencing process, as illustrated *infra*.

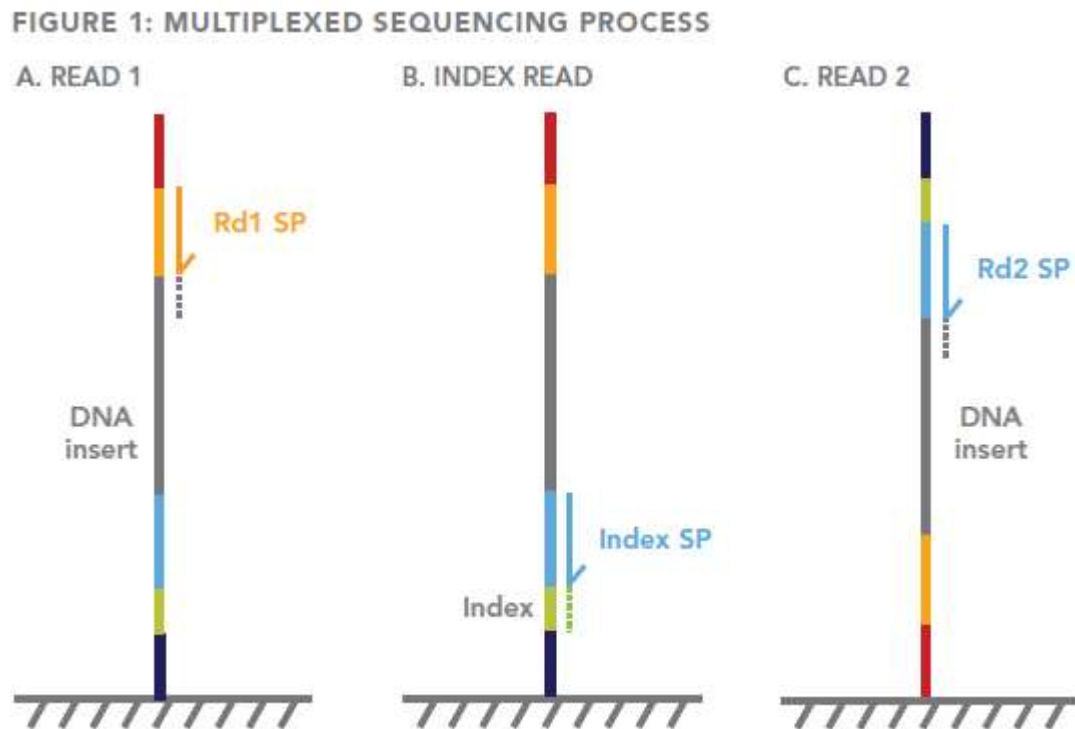


Figure 1 of Illumina (2008) explains “[s]ample multiplexing involves a total of three sequencing reads, *including a separate index read, which is generated*

removed, *only one read with highest mapping quality is retained*. After filtering, a filtered set of deduplicated reads can be used in any downstream analysis.”) (emphases added).

¹⁰⁶ See Ex. 1008, Illumina (2008) at p. 1.

¹⁰⁷ *Id.* at p. 4.

*automatically on the Genome Analyzer equipped with the Paired-End Module.*¹⁰⁸ For the left-hand side of the figure *supra*, that is “A: Read 1,” Illumina (2008) explains that “A: Application read 1 (dotted line) is generated using the Read 1 Sequencing Primer (Rd1 SP).”¹⁰⁹ For the middle section of the figure *supra*, that is “B: Index Read,” Illumina (2008) explains that “B: The read 1 product is removed and the *Index Sequencing Primer (Index SP) is annealed to the same strand to produce the 6-bp index read (dotted line).*”¹¹⁰ For the right-hand side of the figure *supra*, that is “C: Read 2,” Illumina (2008) explains that “C: If a paired-end read is required, the original template strand is used to regenerate the complementary strand.”¹¹¹

F. Metzker (2010)

71. Metzker (2010) is an article I authored that was published in the journal *Nature Reviews Genetics* and is titled, *Sequencing technologies - the next generation*.¹¹² Metzker (2010) was published online on December 8, 2009¹¹³ and is “a technical review of template preparation, sequencing and imaging, genome

¹⁰⁸ *Id.* at Figure 1 legend.

¹⁰⁹ *Id.*

¹¹⁰ *Id.* (emphasis added).

¹¹¹ *Id.*

¹¹² *See* Ex. 1010, Metzker (2010) at 31.

¹¹³ *Id.*

alignment and assembly approaches, and recent advances in current and near-term commercially available NGS instruments.”¹¹⁴ My review article focused on technologies from “Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences, the Polonator instrument and the near-term technology of Pacific Biosciences.”¹¹⁵ Metzker (2010) provide a comprehensive review of “template preparation, sequencing and imaging, genome alignment and assembly, and current NGS platform performance to provide guidance on how these technologies work and how they may be applied to important biological questions.”¹¹⁶

72. Metzker (2010) discusses several aspects of the Illumina/Solexa platform. With respect to template preparation, Metzker (2010) explains that the Illumina/Solexa system uses solid-phase amplification to “produce randomly distributed, clonally amplified clusters from fragment or mate-pair templates on a glass slide.”¹¹⁷ Metzker (2010) further explains that “[h]igh-density forward and reverse primers are covalently attached to the slide, and the ratio of the primers to the template on the support defines the surface density of the amplified clusters. Solid-phase amplification can produce 100–200 million spatially separated

¹¹⁴ *Id.* at Abstract.

¹¹⁵ *Id.* at 31.

¹¹⁶ *Id.*

¹¹⁷ *Id.* at 32.

template clusters (Illumina/Solexa), providing free ends to which a universal sequencing primer can be hybridized to initiate the NGS reaction.”¹¹⁸ Figure 1b of Metzker (2010) demonstrates this process, as illustrated *infra*.

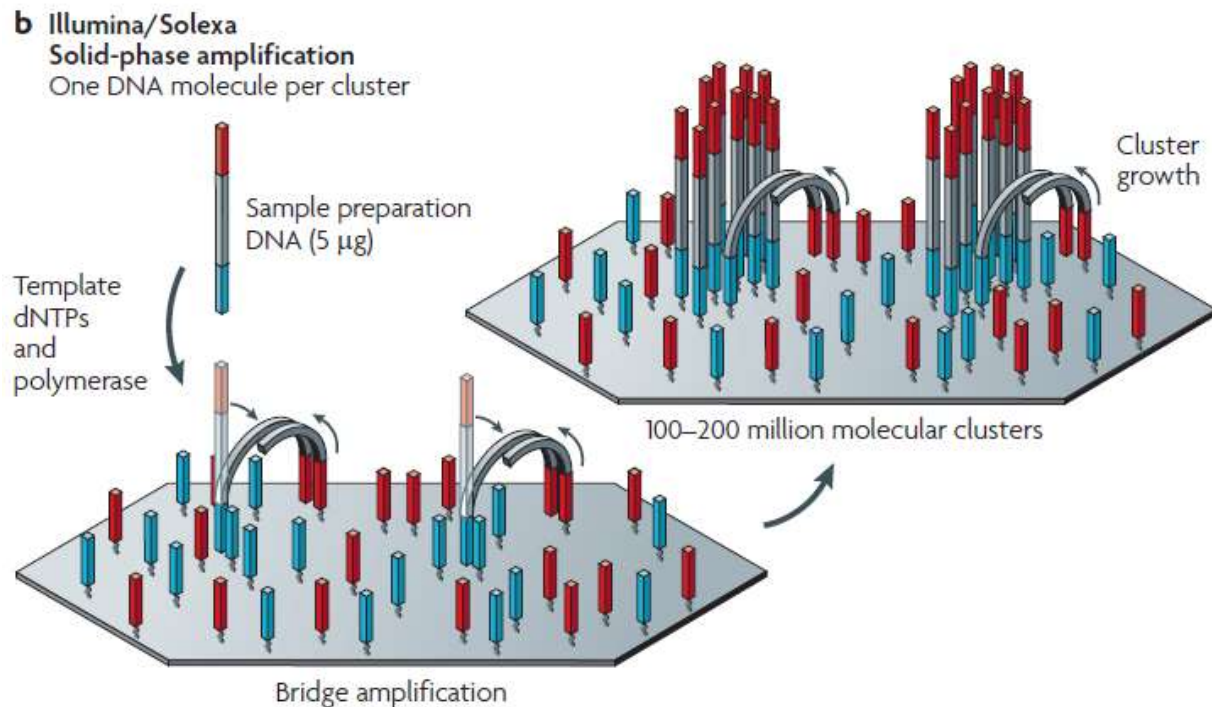


Figure 1b of Metzker (2010) explains that “[s]olid-phase amplification (b) is composed of two basic steps: *initial priming and extending* of the single-stranded, single-molecule template, and *bridge amplification* of the immobilized template with immediately adjacent primers to form clusters.”¹¹⁹ These bridge amplified template formed “molecular clusters” which then sequenced using the Illumina SBS

¹¹⁸ *Id.*

¹¹⁹ *Id.* at Figure 1 legend (emphasis added).

chemistry to generate sequence reads,¹²⁰ that could then be analyzed using SAMtools to identify duplicate sequence reads.

X. ASSERTED GROUNDS

A. Ground 1: Claims 1-3, 5-7, 9-10, and 13-16 Are Anticipated By the '640 Patent Pub. (McCloskey)

73. In my opinion, the '640 Patent Pub. (McCloskey) anticipates Claims 1-3, 5-7, 9-10, and 13-16 of the '241 Patent. Below is a table matching the claim terms in the '241 Patent to the corresponding disclosure in the '640 Patent Pub. (McCloskey).

<u>'241 Patent</u>	<u>'640 Patent Pub. (McCloskey)</u>
"identifier site"	"barcode"
"indexing site"	"batch-stamp"

1. Independent Claim 1 of the '241 Patent

(a) Limitation 1.p

74. Limitation 1.p recites a "method for detecting duplicate sequencing reads, the method comprising." In my opinion, the '640 Patent Pub. (McCloskey) discloses this limitation. The '640 Patent Pub. (McCloskey) teaches the use of barcodes to identify redundant sequences, *i.e.*, duplicate sequences. For example,

¹²⁰ *Id.* at 34-36, Figure 2a.

the '640 Patent Pub. (McCloskey) discloses that “the present invention permits identification of valid sequences, and distinguishes the valid sequences from contaminants and redundant sequences arising from template re-cloning.”¹²¹ As described *supra*, one of ordinary skill in the art would have understood that redundant sequences are duplicate sequences. “***Each genomic fragment is marked prior to amplification, allowing us to identify contaminant and redundant sequences*** and to quantify accurately the proportion of cells carrying a particular sequence variant by counting only distinctly tagged sequences.”¹²²

75. The '640 Patent Pub. (McCloskey) also teaches that sequences having identical target and barcode sequences are redundant. “Redundant hypermethylated sequences recovered from independent bacterial colonies (SEQ ID NO:4 and SEQ ID NO:5), with identical barcodes and methylation patterns.”¹²³ The '640 Patent Pub. (McCloskey) also explains that counting one of two sequences would require discarding or removing the other sequence. “A redundant sequence was also recovered, meaning that a MLM1-positive sequence with an already seen bar-code was recovered. ***One of these two sequences was counted as a valid sequence.***”¹²⁴

¹²¹ See Ex. 1005, '640 Patent Pub. (McCloskey) at [0005].

¹²² *Id.* at [0093] (emphasis added).

¹²³ *Id.* at [0016] [referring to Figure 4B and 4C in the '640 Patent Pub. (McCloskey)].

¹²⁴ *Id.* at [0063] (emphasis added).

The '640 Patent Pub. (McCloskey) further discloses several experiments utilizing barcodes where redundant sequences were observed. “132 sequences were obtained from this experiment. There were 22 sequences with a barcode that was identical to a sequence already obtained (i.e., redundant sequences). . . . The ability to distinguish redundant from valid data is an immensely powerful tool.”¹²⁵

(b) Limitation 1.a

76. Limitation 1.a recites “obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this limitation. For example, the '640 Patent Pub. (McCloskey) teaches two adaptor embodiments, both of which meet Limitation 1.a. One embodiment, the appended adaptor is a hairpin linker containing a barcode and a batch-stamp that is ligated onto a target nucleic acid sequence.¹²⁶

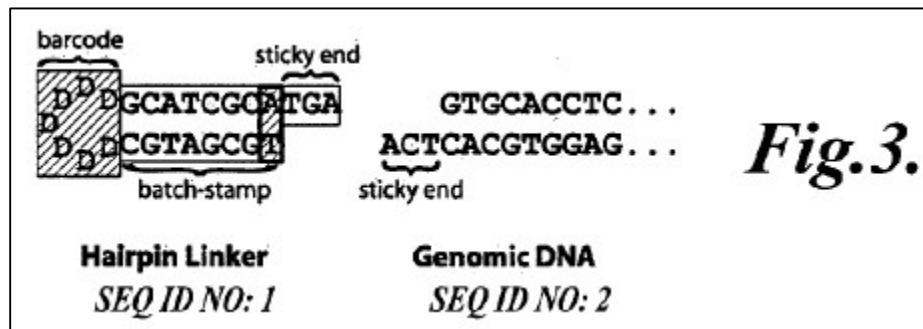
In some embodiments, *the methods comprise the steps of (a) ligating a hairpin linker to a double-stranded target DNA molecule, wherein the hairpin linker comprises a first sequence, a second sequence, and a third sequence, wherein the hairpin linker comprises a first sequence providing experimental identification information, a second sequence providing a random barcode comprising nucleotides selected from the group consisting of adenosines, guanidines [sic], and thymidines, and a third sequence complementary to the first*

¹²⁵ *Id.* at [0053] (emphasis added); *see also id.* at [0046], [0057], [0063], [0068], [0093], [0100], [0103], [0106].

¹²⁶ *Id.* at [0010], [0036], [0095].

sequence; (b) treating the ligated target DNA molecule of step (a) under Suitable conditions to convert cytosines in the ligated target DNA molecule to uracils; (c) amplifying the treated ligated target DNA molecule of step (b) to produce an amplification product; ¹²⁷

The hairpin linker, which is the appended adaptor, is illustrated *infra* in Figure 3 in the '640 Patent Pub. (McCloskey).



The '640 Patent Pub. (McCloskey) explains that “*genomic DNA fragments were labeled with molecular sequence barcodes* and ‘batchstamps’ prior to PCR amplification by *including these molecular labels in the hairpin linker sequence (FIG. 3)* that is used in hairpin-bisulfite PCR.”¹²⁸ The '640 Patent Pub. (McCloskey) also explains that the resulting product is amplified by PCR. “*All PCR products* were analyzed by agarose gel electrophoresis; further cloning and sequencing of appropriately sized products was with TOPO TA Cloning Kits

¹²⁷ *Id.* at [0036] (emphases added). One of ordinary skill in the art would have understood that the '640 Patent Pub. (McCloskey) intended to use the term “guanosine,” which is a nucleoside instead of the term “guanidine.”

¹²⁸ *Id.* at [0093] (emphases added); *see also* [0095]-[0096].

(Invitrogen Life Technologies);. . .”¹²⁹ as such, the PCR amplified products of the hairpin-linker adaptor product satisfies the claimed amplicons.

77. The ’640 Patent Pub. (McCloskey) additionally discloses an appended adaptor as a barcoded oligonucleotide embodiment that satisfies this limitation. For example, the ’640 Patent Pub. (McCloskey) teaches a method utilizing the appended adaptor that comprises the following steps.

Thus, some embodiment of the methods of the invention comprise the steps of “(a) ***contacting a target nucleic acid molecule in a sample with a bar-coded oligo nucleotide under suitable conditions to anneal the bar-coded oligonucleotide to the target nucleic acid molecule***, wherein the bar-coded oligonucleotide comprises a first sequence complementary to the target nucleic acid molecule, ***a second sequence providing a random barcode***, a third sequence that is not complementary to any sequence in the sample, and a fourth sequence providing experimental identification information, and wherein the 5’ end of the bar-coded oligonucleotide comprises a tethering sequence that is complementary to a sequence 5’ to the first sequence; (b) ***extending the annealed bar-coded oligonucleotide*** using the target nucleic acid molecule as a template to produce a bar-coded target nucleic acid molecule; (c) ***amplifying the bar-coded target nucleic acid*** molecule using a primer that binds to the third sequence to produce an amplification product; . . .”¹³⁰

This process of appended adaptor to the nucleic acid and amplifying the product is illustrated *infra* in Figures 2B-2D in the ’640 Patent Pub. (McCloskey).

¹²⁹ *Id.* at [0098] (emphasis added); *see also* [0097] for the PCR conditions.

¹³⁰ *Id.* at [0035] (emphases added).

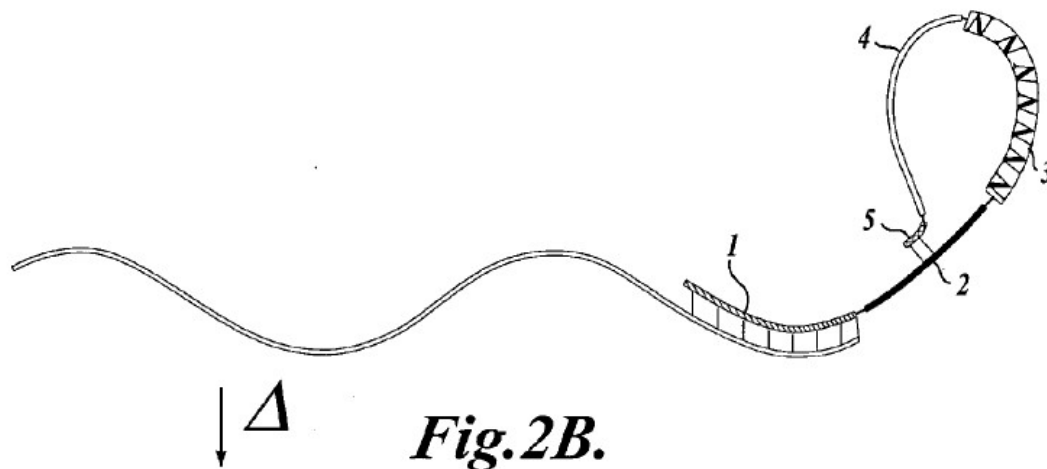


Fig.2B.

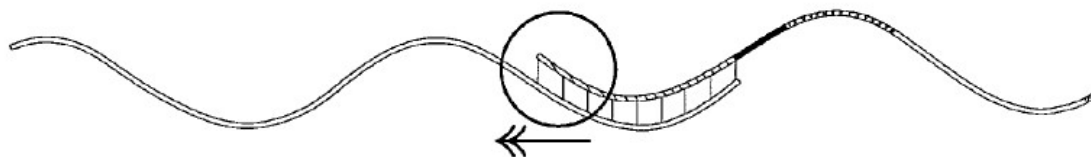


Fig.2C.

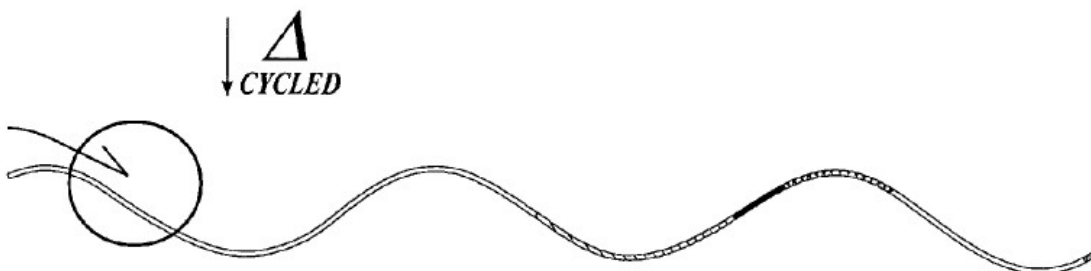


Fig.2D.

Figure 1 in the '640 Patent Pub. (McCloskey) shows the “random barcode (3)” as well as the “batch-stamp sequence (2).”¹³¹ The '640 Patent Pub. (McCloskey) explains the steps of Figure 2A-2D.

FIGS. 2A-2D shows the steps in a representative method of the invention for authenticating a DNA amplification product. A: A target DNA molecule is denatured; B: *the bar-coded oligonucleotide shown in FIG. 1 is annealed to one strand of the target DNA molecule*; C: *the bar-coded oligonucleotide is extended* using Sequenase to synthesize a bar-coded target DNA molecule; D: *the bar-coded target DNA molecule is amplified* using the polymerase chain reaction (PCR). The closed circles represent polymerase molecules.”¹³²

The amplification products resulting from this process satisfies the claims amplicons. This embodiment is also used in Examples 1-5 of the '640 Patent Pub. (McCloskey).¹³³

(c) Limitation 1.b

78. Limitation 1.b recites “wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this limitation. For example, the '640 Patent Pub. (McCloskey)'s barcode, which is incorporated into both the hairpin linker adaptor and barcoded oligonucleotide adaptor embodiments, as

¹³¹ *Id.* at [0013].

¹³² *Id.* at [0014] (emphasis added).

¹³³ *Id.* at [0043]-[0091].

described *supra* at Limitation 1.a, that satisfies the claimed “identifier site.” The ’640 Patent Pub. (McCloskey) explains that its “random barcode,” which comprises “nucleotides selected from the group consisting of adenosines, guanidines, [*sic*] and thymidines.”¹³⁴ The ’640 Patent Pub. (McCloskey) also explains that its “random barcode” “refers to an arbitrary sequence that can uniquely identify a target nucleic acid in an experiment.”¹³⁵ This random barcode sequence should be of sufficient length to provide “a unique identity to each target nucleic acid molecule in the sample prior to amplification” and can range “between 3 and 30 nucleotides.”¹³⁶

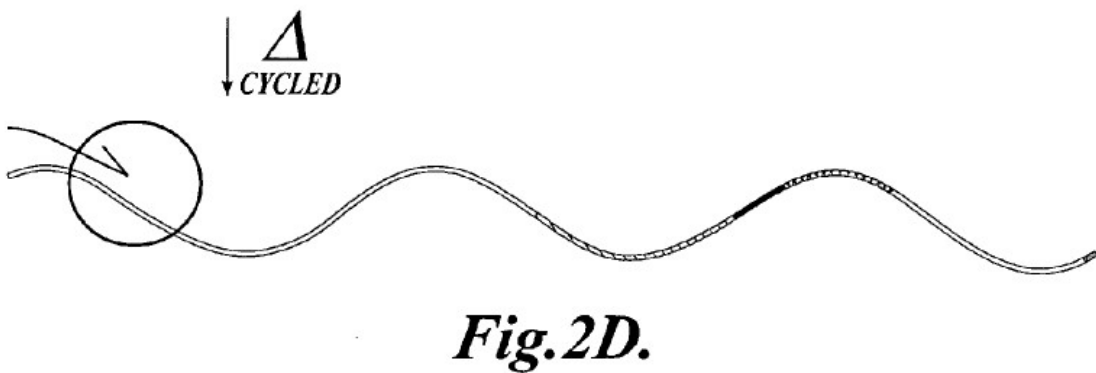
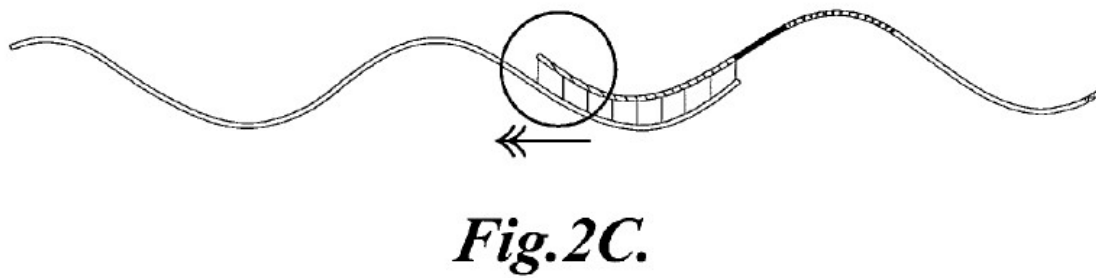
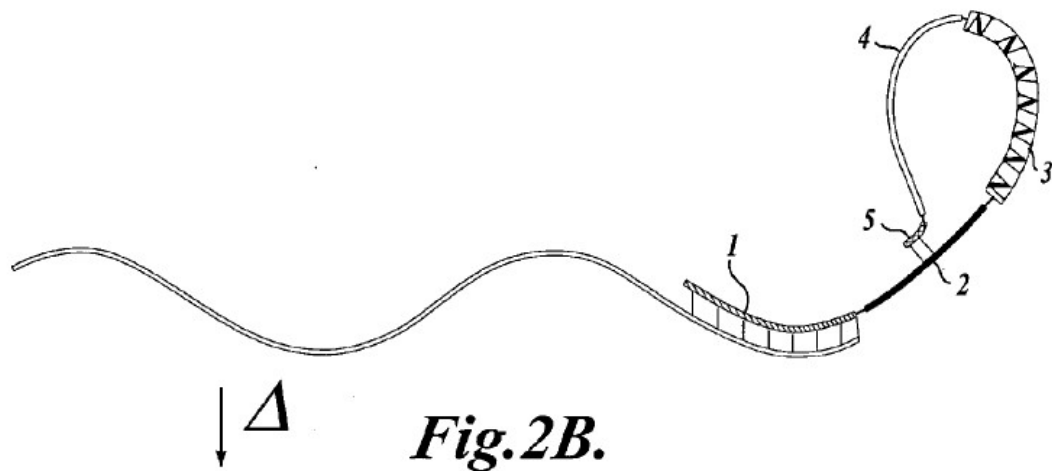
79. The ’640 Patent Pub. (McCloskey) provides several examples where a barcode sequence is appended to a target DNA sequence and the resulting molecule is subsequently amplified using PCR. For instance, the oligonucleotide adaptor embodiment, which is shown in Example 2, includes “a 7-nucleotide random barcode,” which after an annealing and extension step, is amplified by PCR.¹³⁷ Figures 2A-D of the ’640 Patent Pub. (McCloskey) illustrates process of having a plurality of nucleotides unique to the amplified fragment.

¹³⁴ *Id.* at [0010]. One of ordinary skill in the art would have understood that the ’640 Patent Pub. (McCloskey) intended to use the term “guanosine,” which is a nucleoside instead of the term “guanidine.”

¹³⁵ *Id.* at [0021].

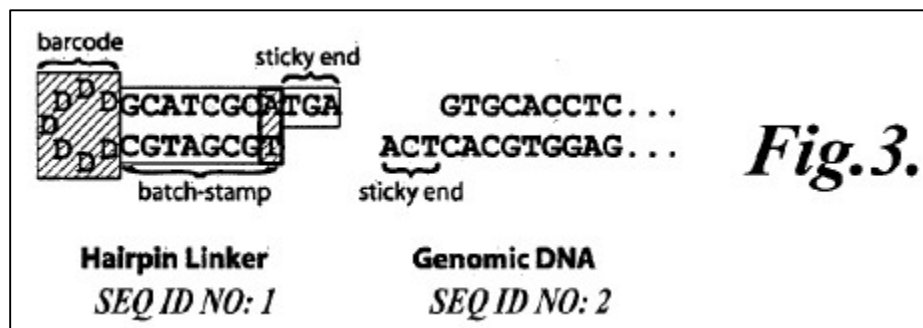
¹³⁶ *Id.*

¹³⁷ *Id.* at [0057].



The random barcode in Figure 2B is labelled as sequence “3” that comprises seven “N” nucleotide residues, which would be unique for amplicon generated in Figure 2D.

80. The random barcode is also described in Example 6 of the '640 Patent Pub. (McCloskey), in which it is included in a hairpin linker adaptor that is ligated to a target DNA sequence.¹³⁸ This hairpin linker product is subsequently amplified by PCR and then sequenced.¹³⁹ Figure 3 of the '640 Patent Pub. (McCloskey) shows its random barcode adaptor, as illustrated *infra*.



The seven “D” residues “represents a nucleotide randomly selected from A, G, and T.”¹⁴⁰ The '640 Patent Pub. (McCloskey) explains that “FIG. 3 shows a schematic representation of a bar-coded and batch-stamped hairpin linker (SEQ ID NO: 1), designed for ligation to DraIII-cut genomic DNA of FMR1 (SEQ ID NO:2), as described in EXAMPLE 6.”¹⁴¹ Other examples in the '640 Patent Pub. (McCloskey) similarly disclose barcoded DNA products.¹⁴² The random barcode

¹³⁸ *Id.* at [0095], [0097]-[0098].

¹³⁹ *Id.* at [0095], [0097]-[0098].

¹⁴⁰ *Id.* at [0015].

¹⁴¹ *Id.*

¹⁴² *Id.* at [0046]-[0047], [0057]-[0060], [0068]-[0073], [0076], [0081]-[0084], [0095]-[0098], [0100]-[0104].

that is included in both the barcoded oligonucleotide adaptor and in the hairpin linker adaptor satisfies the claimed identifier site.

(d) Limitation 1.c

81. Limitation 1.c recites “sequencing the amplicons to generate sequence reads that include identifier and target sequences.” In my opinion, the ’640 Patent Pub. (McCloskey) discloses this limitation. The ’640 Patent Pub. (McCloskey) teaches that the bar-coded DNA fragment products resulting from both the barcoded oligonucleotide and hairpin linker embodiments described *supra* in Limitation 1.a and 1.b are sequenced.

82. For example, the ’640 Patent Pub. (McCloskey) teaches that the amplified bar-coded DNA fragments of the oligonucleotide embodiment are sequenced.

Thus, some embodiments of the methods of the invention comprise the steps of (a) contacting a target nucleic acid molecule in a sample with a bar-coded oligo nucleotide under suitable conditions to anneal the bar-coded oligonucleotide to the target nucleic acid molecule, wherein the bar-coded oligonucleotide comprises a first sequence complementary to the target nucleic acid molecule, a second sequence providing a random barcode, a third sequence that is not complementary to any sequence in the sample, and a fourth sequence providing experimental identification information, and wherein the 5’ end of the bar-coded oligonucleotide comprises a tethering sequence that is complementary to a sequence 5’ to the first sequence; (b) extending the annealed bar-coded oligonucleotide using the target nucleic acid molecule as a template to produce a bar-coded target nucleic acid molecule; (c) amplifying the bar-coded target nucleic acid molecule using a primer that binds to the third

sequence to produce an amplification product; and (d) *authenticating the amplification product by detecting the presence of the second sequence and the fourth sequence in the amplification product.*”¹⁴³

83. The ’640 Patent Pub. (McCloskey) further explains that its sequencing methods were known in the art. “*Methods and conditions for analyzing nucleic acid molecules by gel electrophoresis, cloning, and sequencing are standard in the art. Exemplary conditions for analyzing bar-coded target nucleic acid molecules by gel electrophoresis, cloning, and sequencing are described in EXAMPLES 1-4 and 6.*”¹⁴⁴

84. The amplified products of the oligonucleotide embodiments are also shown to be sequenced in Examples 1-4.¹⁴⁵ For example, the ’640 Patent Pub. (McCloskey) explains that “132 sequences were obtained from this experiment. There were 22 sequences with a barcode that was identical to a sequence already obtained (i.e., redundant sequences).”¹⁴⁶ The ’640 Patent Pub. (McCloskey) also explains that “46 sequences were obtained from this experiment. The barcodes, batch-stamps, and the annealing temperature of four exemplary bar-coded

¹⁴³ *Id.* at [0035] (emphasis added). One of ordinary skill in the art would have understood that “authenticating the amplification product” would require sequencing the amplified product.

¹⁴⁴ *Id.* at [0034] (emphases added).

¹⁴⁵ *Id.* at [0043]-[0087].

¹⁴⁶ *Id.* at [0053].

sequences are shown in TABLE 1. 36 of the 46 were desirable and valid sequences, each of which had a unique barcode.”¹⁴⁷ The ’640 Patent Pub. (McCloskey) further explains that “107 sequences were obtained from this experiment.”¹⁴⁸

85. The barcoded DNA amplification products of the hairpin linker embodiment are sequenced as well.

In some embodiments, the methods comprise the steps of (a) ligating a hairpin linker to a double-stranded target DNA molecule, wherein the hairpin linker comprises a first sequence, a second sequence, and a third sequence, wherein the hairpin linker comprises a first sequence providing experimental identification information, a second sequence providing a random barcode comprising nucleotides selected from the group consisting of adenosines, guanidines [*sic*], and thymidines, and a third sequence complementary to the first sequence; (b) treating the ligated target DNA molecule of step (a) under Suitable conditions to convert cytosines in the ligated target DNA molecule to uracils; (c) amplifying the treated ligated target DNA molecule of step (b) to produce an amplification product; and (d) ***authenticating the amplification product of step (c) by detecting the presence of the second sequence in the amplification product.***¹⁴⁹

¹⁴⁷ *Id.* at [0063].

¹⁴⁸ *Id.* at [0076].

¹⁴⁹ *Id.* at [0036] (emphasis added). One of ordinary skill in the art would have understood that “authenticating the amplification product” would require sequencing the amplified product. One of ordinary skill in the art would have also understood that the ’640 Patent Pub. (McCloskey) intended to use the term “guanosine,” which is a nucleoside instead of the term “guanidine.”

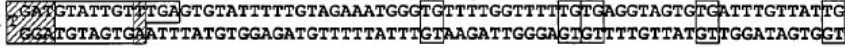
86. The amplified products of the hairpin linker embodiment are also described as being sequenced in Example 6.¹⁵⁰ For example, the '640 Patent Pub. (McCloskey) explains that different classes of sequences were obtained in Example 6.

*The classes of sequences recovered included hypermethylated sequences with distinctive barcodes and patterns of methylation (FIG. 4A), redundant hypermethylated sequences with identical barcodes and methylation patterns (FIGS. 4B and 4C), hypomethylated sequences with distinctive barcodes (FIG. 4D), redundant hypomethylated sequences with identical barcodes (FIGS. 4E and 4F), and contaminant sequences with an original linker that predates the barcoding (FIG. 4G)."*¹⁵¹

Figure 4 of the '640 Patent Pub. (McCloskey), illustrated *infra*, additionally shows the different classes of sequence reads obtained from the barcoded DNA amplification products.

¹⁵⁰ *Id.* at [0098]-[0107].

¹⁵¹ *Id.* at [0103] (emphases added).

Fig.4A.		<i>SEQ ID NO: 3</i>
Fig.4B.	* 	<i>SEQ ID NO: 4</i>
Fig.4C.	* 	<i>SEQ ID NO: 5</i>
Fig.4D.		<i>SEQ ID NO: 6</i>
Fig.4E.	+ 	<i>SEQ ID NO: 7</i>
Fig.4F.	+ 	<i>SEQ ID NO: 8</i>
Fig.4G.		<i>SEQ ID NO: 9</i>

(e) Limitation 1.d

87. Limitation 1.d recites “identifying sequence reads with identical identifier and target sequences as duplicates.” In my opinion, the ’640 Patent Pub. (McCloskey) discloses this limitation. The ’640 Patent Pub. (McCloskey) teaches that identical sequence reads having the same barcode are redundant, or in other words, duplicates. The ’640 Patent Pub. (McCloskey) applies this analysis to the sequence reads from both the barcoded oligonucleotide and hairpin linker embodiments.

88. For instance, with respect to the oligonucleotide embodiment disclosed in Example 1, the ’640 Patent Pub. (McCloskey) explains that “132 sequences were obtained from this experiment. *There were 22 sequences with a barcode that was identical to a sequence already obtained (i.e., redundant*

sequences).”¹⁵² The ’640 Patent Pub. (McCloskey) also explains in Example 2 that duplicate sequence reads were identified having identical identifier and target sequences.

46 sequences were obtained from this experiment. The barcodes, batch-stamps, and the annealing temperature of four exemplary bar-coded sequences are shown in TABLE 1. 36 of the 46 were desirable and valid sequences, each of which had a unique barcode. Only one of the 46 sequences was a contaminant sequence: it contained the batch-stamp of the HBP oligo used in EXAMPLE 1 and a 5-nucleotide barcode (MLM1 43 in Table 1). ***A redundant sequence was also recovered, meaning that a MLM1-positive sequence with an already seen bar-code was recovered.***¹⁵³

89. Identification of a duplicate sequence read was also described in Example 6 in the ’640 Patent Pub. (McCloskey), which involves the hairpin linker embodiment. For example, the ’640 Patent Pub. (McCloskey) explains in Example 6 that different classes of sequences identified were “***redundant hypermethylated sequences with identical barcodes and methylation patterns*** (FIGS. 4B and 4C)” and “***redundant hypomethylated sequences with identical barcodes*** (FIGS. 4E and 4F).”¹⁵⁴

90. These disclosures in the ’640 Patent Pub. (McCloskey) demonstrate that an identical target sequence (*e.g.*, “redundant hypermethylated sequences” and

¹⁵² *Id.* at [0053] (emphasis added).

¹⁵³ *Id.* at [0063] (emphasis added).

¹⁵⁴ *Id.* at [0103] (emphases added).

“redundant hypomethylated sequences”) and an identical barcode are used to identify which sequences were duplicates.

2. Claim 2 of the '241 Patent

91. Claim 2 recites “[t]he method of claim 1, further comprising removing a duplicate read from the sequence reads.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. For example, the '640 Patent Pub. (McCloskey)'s Example 2 resulted in the identification of two identical sequences, where both sequences were tagged with the same barcode. “***A redundant sequence was also recovered***, meaning that a MLM1-positive sequence with an already seen bar-code was recovered.”¹⁵⁵ The '640 Patent Pub. (McCloskey) further disclosed that one of the two redundant sequences were removed from the analysis: “***One of these two sequences was counted as a valid sequence.***”¹⁵⁶ As described *supra*, one of ordinary skill in the art would have understood that counting one of two sequences would require discarding or removing the other sequence. Thus, the '640 Patent Pub. (McCloskey)'s description of the removal of a redundant (*i.e.*, duplicate) sequence read satisfies this claim.

¹⁵⁵ *Id.* at [0063] (emphasis added).

¹⁵⁶ *Id.* (emphasis added).

3. Claim 3 of the '241 Patent

92. Claim 3 recites “[t]he method of claim 1, wherein the identifier site is sequenced with the target sequence.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Limitation 1.a, the '640 Patent Pub. (McCloskey) discloses two embodiments of barcoded DNA amplification products: one resulting from ligation of a hairpin adaptor to the target sequence, and another resulting from extension of a barcoded oligonucleotide adaptor. In both scenarios, the barcode is sequenced with the target DNA sequence.

93. For example, Figures 4A-G in the '640 Patent Pub. (McCloskey) show that the barcode sequence in the hairpin adaptor is sequenced with the target DNA sequence, as illustrated *infra*.



The '640 Patent Pub. (McCloskey) further explains that the barcodes are sequenced with the target sequence.

FIGS. 4A-4G show FMR1 promoter sequences, with inferred methylation states of CpG sites, recovered from male fragile X patients using hairpin-bisulfite PCR with linker bar-coding and batch-stamping, as described in EXAMPLE 6. . . . All sequences show 100% conversion of non-CpG cytosines. A: A distinctive hypermethylated sequence (SEQ ID NO:3); B, C: **Redundant hypermethylated sequences recovered** from independent bacterial colonies (SEQ ID NO:4 and SEQ ID NO:5), **with identical barcodes and methylation patterns**; D: A hypomethylated sequence with a distinctive barcode (SEQ ID NO:6); E, F: **Redundant hypomethylated sequences with identical barcodes** recovered from independent bacterial colonies (SEQ ID NO:7 and SEQ ID NO:8)."¹⁵⁷

94. The barcode sequence and target sequenced are sequenced together in the barcoded oligonucleotide embodiment as well. Because the barcoded oligonucleotide embodiment contains a "leftward primer binding site (4)," both sequences are sequenced in a single run.¹⁵⁸ The '640 Patent Pub. (McCloskey) further explains that both sequences are read together. "132 sequences were obtained from this experiment. **There were 22 sequences with a barcode that was identical to a sequence already obtained** (i.e., redundant sequences)."¹⁵⁹ Thus, the

¹⁵⁷ *Id.* at [0016] (emphases added).

¹⁵⁸ *Id.* at [0013].

¹⁵⁹ *Id.* at [0053] (emphasis added); *see also id.* at [0063]. ("46 sequences were obtained from this experiment. The barcodes, batch-stamps, and the annealing temperature of four exemplary bar-coded sequences are shown in TABLE 1. **36 of**

'640 Patent Pub. (McCloskey)'s description that the identifier site is sequenced with the target sequence satisfies the claim.

4. Claim 5 of the '241 Patent

95. Claim 5 recites “[t]he method of claim 1, wherein the amplicons include amplified fragments of cDNA.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. For example, the '640 Patent Pub. (McCloskey) discloses that its methods are applicable to RNA sequences, and one of ordinary skill in the art would have known that applying the '640 Patent Pub. (McCloskey)'s methods to RNA would first require reverse transcribing the RNA sequence into cDNA sequence, which ultimately results in cDNA amplicons. “The nucleotides can be naturally-occurring, artificial and/or modified nucleotides, or any other complementary subunits. Thus, *the target nucleic acid molecule may be any nucleic acid molecule, including, but not limited to DNA and RNA (such as ribosomal RNA, messenger RNA, and small untranslated RNA)*.”¹⁶⁰

96. Example 5 in the '640 Patent Pub. (McCloskey) also discloses barcoding a target RNA molecule. “*The design of barcode oligonucleotides for barcoding a target RNA molecule* will encompass the same concepts as described

the 46 were desirable and valid sequences, each of which had a unique barcode.”) (emphasis added).

¹⁶⁰ *Id.* at [0018] (emphases added).

in EXAMPLES 1-4 one [*sic*]. The barcoded oligonucleotide contains a region complementary to the target RNA strand. The bar-coded oligonucleotide also contains a batch-stamp region, a barcode region, and a unique primer binding site, as described in EXAMPLES 1-4.”¹⁶¹

97. The '640 Patent Pub. (McCloskey) further explains that “[t]he ***bar-coded oligonucleotide is used in a reverse transcriptase reaction as is standard in the art***. After one round of reverse transcription, the reaction is stopped and the mixture is digested with an appropriate amount of uracil glycosylase to destroy any remaining RNA template (which includes uracil) in the mix.”¹⁶² As described *supra*, one of ordinary skill in the art would have understood that *a reverse transcriptase reaction* converts the RNA molecule is a barcoded cDNA molecule. The subsequent PCR amplification process¹⁶³ produces cDNA amplicons. Thus, the '640 Patent Pub. (McCloskey)'s description of reverse transcribing RNA into cDNA satisfies the claim.

5. Claim 6 of the '241 Patent

98. Claim 6 recites “[t]he method of claim 1, wherein the amplicons include amplified fragments of genomic DNA.” In my opinion, the '640 Patent

¹⁶¹ *Id.* at [0089] (emphasis added).

¹⁶² *Id.* at [0090] (emphasis added).

¹⁶³ *Id.* (“Post column-purification ***standard PCR is carried out with cycles 2-35 cycles.***”) (emphasis added).

Pub. (McCloskey) discloses this claim. For example, the '640 Patent Pub. (McCloskey) teaches that its methods are applicable to genomic DNA and provides several examples of this. For instance, the '640 Patent Pub. (McCloskey) explains that “[i]n the practice of the present invention *nucleic acid molecules (e.g., genomic DNA fragments) are labeled with distinct sequence tags* prior to PCR amplification.”¹⁶⁴ The '640 Patent Pub. (McCloskey)'s examples use genomic DNA as well. For example, the '640 Patent Pub. (McCloskey) explains for Example 6 that “FIG. 3 shows a schematic representation of a bar-coded and batch-stamped hairpin linker (SEQ ID NO: 1), designed for ligation to *DraIII-cut genomic DNA of FMR1* (SEQ ID NO:2), as described in EXAMPLE 6.”¹⁶⁵ The '640 Patent Pub. (McCloskey) also explains that “Target DNA: *Human genomic DNA* was isolated from blood using a standard protocol.”¹⁶⁶ And in again in Example, 6, the '640 Patent Pub. (McCloskey) explains that “*5 μg of genomic DNA* was cleaved by 10 U each of restriction endonucleases *DraIII* and *AluI* for 1 h at 37° C.”¹⁶⁷ Thus, the '640 Patent Pub. (McCloskey)'s description of a genomic DNA satisfies the claim.

¹⁶⁴ *Id.* at [0005] (emphasis added).

¹⁶⁵ *Id.* at [0015] (emphasis added).

¹⁶⁶ *Id.* at [0045] (emphasis added).

¹⁶⁷ *Id.* at [0095] (emphasis added).

6. Claim 7 of the '241 Patent

99. Claim 7 recites “[t]he method of claim 1, wherein each adaptor includes a primer binding site.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Limitation 1.a, the '640 Patent Pub. (McCloskey) teaches an oligonucleotide adaptor embodiment. This oligonucleotide adaptor also contains a “leftward primer binding side,” which satisfies this claim.¹⁶⁸ Thus, the '640 Patent Pub. (McCloskey)’s description of a primer binding site satisfies the claim.

7. Claim 9 of the '241 Patent

100. Claim 9 recites “[t]he method of claim 1, wherein each unique identifier is at least 6 nucleotides in length.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Limitations 1.a and 1.b, the '640 Patent Pub. (McCloskey) teaches a “barcode” that satisfies the claimed “unique identifier.” The '640 Patent Pub. (McCloskey) explains that this barcode, also referred to as “the second sequence,”¹⁶⁹ may be of varying lengths, including at least 6 nucleotides in length.

The length of the second sequence is sufficient to provide, with high probability, a unique identity to each target nucleic acid molecule in the sample prior to amplification. For example, ***a second sequence of 7 random nucleotides N selected from A, G, C, and T*** will provide a maximum of 4^7 or 16,384 unique

¹⁶⁸ *Id.* at [0013]; *see also id.* at Figures 1, 2B-2D.

¹⁶⁹ *Id.* at [0021].

barcodes. In some embodiments, *the length of the second sequence is between 3 and 30 nucleotides, such as between 5 and 25 nucleotides or between 7 and 13 nucleotides.*¹⁷⁰

Thus, the '640 Patent Pub. (McCloskey)'s description of the length of the unique identifier can be as long as 30 nucleotides satisfies the claim.

8. Claim 10 of the '241 Patent

101. Claim 10 recites “[t]he method of claim 1, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Limitations 1.a and 1.b, the '640 Patent Pub. (McCloskey) discloses a hairpin linker adaptor embodiment and a barcoded oligonucleotide adaptor embodiment. Both of these embodiments satisfy this claim. The '640 Patent Pub. (McCloskey) teaches that the hairpin linker adaptor can be ligated to the target nucleic acid sequence.

In some embodiments, the methods comprise the steps of (a) *ligating a hairpin linker to a double-stranded target DNA molecule*, wherein the hairpin linker comprises a first sequence, a second sequence, and a third sequence, wherein the hairpin linker comprises a first sequence providing experimental identification information, a second sequence providing a random barcode comprising nucleotides selected from the group

¹⁷⁰ *Id.* (emphases added).

consisting of adenosines, guanidines [*sic*], and thymidines, and a third sequence complementary to the first sequence; . . .¹⁷¹

102. The '640 Patent Pub. (McCloskey) also teaches that an barcoded oligonucleotide adaptor can be annealed to a target nucleic acid sequence, the product of which is then extended to produce a barcoded nucleic acid product.

In step (b) of the methods of this aspect of the invention, ***the annealed bar-coded oligonucleotide is extended using the target nucleic acid molecule as a template to produce a bar-coded target nucleic acid molecule.*** Methods for extending an annealed oligonucleotide using a nucleic acid molecule as a template are standard in the art.”¹⁷²

This process of annealing an oligonucleotide following by extension using a polymerase is known as priming reaction to append the adaptors to the fragments, and thus, the '640 Patent Pub. (McCloskey)'s description of a ligation reaction and a primer extension reaction satisfies the claim.

9. Claim 13 of the '241 Patent

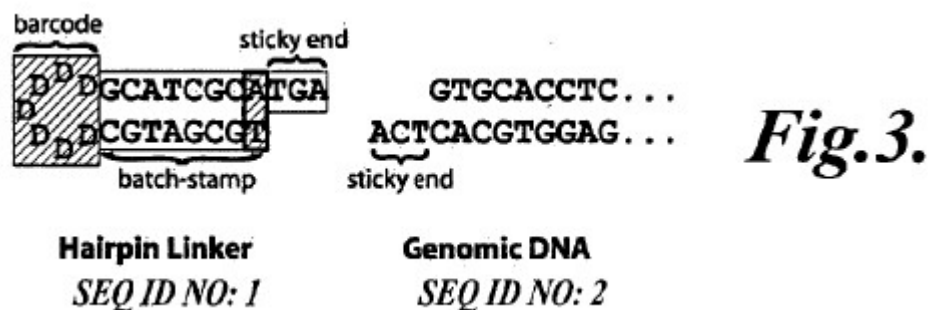
103. Claim 13 recites “[t]he method of claim 1, wherein each adaptor further comprises an indexing site.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Limitations 1.a and 1.b,

¹⁷¹ *Id.* at [0036] (emphases added). One of ordinary skill in the art would have understood that the '640 Patent Pub. (McCloskey) intended to use the term “guanosine,” which is a nucleoside instead of the term “guanidine.” *See also id.* at [0010], [0015], [0037], [0040], [0095], [0100], Figure 3.

¹⁷² *Id.* at [0030] (emphases added); *see also id.* at [0008]-[0009], [0014], [0017], [0032], [0035], [0048], [0058], [0071], [0082], Figure 2C.

the '640 Patent Pub. (McCloskey) discloses a hairpin linker adaptor embodiment and a barcoded oligonucleotide adaptor embodiment. Both of these embodiments include a “batch-stamp,” which satisfies the claimed “indexing site.” The '640 Patent Pub. (McCloskey) explains that a “‘batch-stamp’ refers to information that uniquely identifies a specific experiment. For example, *a batch-stamp may identify a sample*, a patient, and/or an analysis date.”¹⁷³

104. The hairpin linker embodiment includes a batch-stamp sequence. “**FIG. 3 shows a schematic representation of a bar-coded and batch-stamped hairpin linker** (SEQ ID NO:1), designed for ligation to DraIII-cut genomic DNA of FMR1 (SEQ ID NO:2), as described in EXAMPLE 6.”¹⁷⁴ Figure 3 of the '640 Patent Pub. (McCloskey) is reproduced *infra*:



The '640 Patent Pub. (McCloskey) further discloses that its Example 6 method involves the use of a batch-stamp. “Ligation of the hairpin linker (5’ AGC-

¹⁷³ *Id.* at [0027]. (emphasis added).

¹⁷⁴ *Id.* at [0015]. (emphasis added).

GATGCDDDDDDDD GCATCGCTTGA 3', SEQ ID NO:1) *with variations in the non-random nucleotides for batch-stamps*"¹⁷⁵

105. The barcoded oligonucleotide embodiment includes a batch-stamp as well. "FIG. 1 shows *an exemplary bar-coded oligonucleotide of the invention containing* a sequence complementary to the target nuclei acid molecules *(1)*, *a batch-stamp sequence* (2), random barcode (3), a leftward primer binding site (4), and a 5' tethering sequence (5), as described in EXAMPLES 1-5."¹⁷⁶ Thus, the '640 Patent Pub. (McCloskey)'s description of a batch-stamp sequence satisfies the claim.

10. Claim 14 of the '241 Patent

106. This claim recites "[t]he method of claim 1, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length." In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Limitations 1.a and 1.b, the '640 Patent Pub. (McCloskey) teaches a "barcode" that satisfies the claimed "unique identifier." The '640 Patent Pub. (McCloskey) explains that this barcode, also referred to as "the second sequence,"¹⁷⁷ may be of varying lengths, including at least 6 nucleotides in length.

¹⁷⁵ *Id.* at [0095]. (emphasis added).

¹⁷⁶ *Id.* at [0013]. (emphases added); *see also id.* at Examples 1-5.

¹⁷⁷ *Id.* at [0021].

The length of the second sequence is sufficient to provide, with high probability, a unique identity to each target nucleic acid molecule in the sample prior to amplification. For example, *a second sequence of 7 random nucleotides N selected from A, G, C, and T* will provide a maximum of 4^7 or 16,384 unique barcodes. In some embodiments, *the length of the second sequence is between 3 and 30 nucleotides, such as between 5 and 25 nucleotides or between 7 and 13 nucleotides.*¹⁷⁸

Thus, the '640 Patent Pub. (McCloskey)'s description that the unique identifier can be at least 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length satisfies the claim.

11. Claim 15 of the '241 Patent

107. This claim recites “[t]he method of claim 1, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and performing an amplification reaction to obtain the amplicons.” In my opinion, the '640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Limitation 1.a, the '640 Patent Pub. (McCloskey) discloses ligated a hairpin linker adaptor to fragmented DNA. Thus, the '640 Patent Pub. (McCloskey)'s use of a hairpin linker adaptor embodiment satisfies this claim.

108. Thus, the '640 Patent Pub. (McCloskey) explains that, in Example 6, “genomic DNA fragments were labeled with molecular sequence barcodes and ‘batchstamps’ *prior to PCR amplification* by including these molecular labels in

¹⁷⁸ *Id.* (emphases added).

the hairpin linker sequence (FIG. 3) that is used in hairpin-bisulfite PCR. . . . Each genomic fragment is marked *prior to amplification*, allowing us to identify contaminant and redundant sequences and to quantify accurately the proportion of cells carrying a particular sequence variant by counting only distinctly tagged sequences.”¹⁷⁹ Thus, the ’640 Patent Pub. (McCloskey)’s disclosure of an amplification reaction to obtain the amplicons from adaptor appended DNA fragments satisfies this claim.

12. Claim 16 of the ’241 Patent

109. This claim recites “[t]he method of claim 1, wherein the amplicons include amplified fragment of a nucleic acid from a first sample with an appended adaptor that includes a first index and amplified fragment of a nucleic acid from a second sample with an appended adaptor that includes a second index, and the method includes separating sequence reads by sample based on the first or second index.” In my opinion, the ’640 Patent Pub. (McCloskey) discloses this claim. As described *supra* at Claim 13 of the ’241 Patent, the ’640 Patent Pub. (McCloskey)’s batch-stamp satisfies the claimed indexing site.

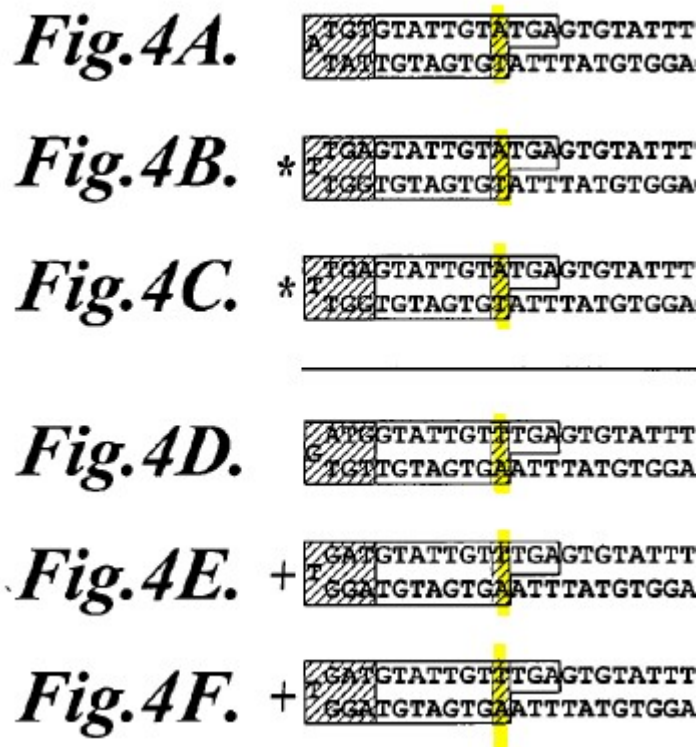
110. The ’640 Patent Pub. (McCloskey) further discloses the analysis of nucleic acids from different samples.

In addition to adding the random barcode, molecules were
“batch-stamped” by encoding the hairpin linker with

¹⁷⁹ *Id.* at [0093] (emphases added).

information that would designate the sample analyzed and the date of analysis. Multiple variants of the hairpin linker were designed by changing nucleotides in the stem of the linker. *These stem changes represented different batches of linkers, each of which were used for the analysis of a different sample.* Thus, the resulting sequences each bear a consistent ‘batch-stamp’ encoded in the stem, and a randomly variable barcode encoded in the loop (FIG. 3).¹⁸⁰

111. The ’640 Patent Pub. (McCloskey) additionally discloses an example where sequences with different batch stamps were identified as originating from different samples. Figures 4A-4F of the ’640 Patent Pub. (McCloskey) demonstrate two different samples containing different batch-stamps, as illustrated *infra*.



¹⁸⁰ *Id.* at [0102] (emphases added).

In this excerpt of Figures 4A-F, I highlighted a portion of the batch-stamp in yellow. In the sequences shown in Figures 4A-C, the highlighted portion contains an “A” on the top strand and a “T” on the bottom strand. In the sequences shown in Figures D-F, the highlighted portion contains a “T” on the top strand and an “A” on the bottom strand. The ’640 Patent Pub. (McCloskey) explains that “[s]equences A to C carry a different batch-stamp than sequences D to F, with the inversion of the A-T base pair, *confirming that these sequence sets came from different DNA samples.*”¹⁸¹

112. One of ordinary skill in the art would have known that nucleic acids bearing different batch-stamps originated from different samples. One of ordinary skill in the art would have further known that the steps of this claim are necessarily performed when identifying different samples. Thus, the ’640 Patent Pub. (McCloskey)’s disclosure of tagging a first and second sample with different batch-stamps and subsequently identifying sequences with different batch-stamps as originating from different samples satisfies this claim.

¹⁸¹ *Id.* at [0016] (emphasis added).

B. Ground II: Claims 1-3, 5-11, 13-16 Are Anticipated by '202 Patent Pub. (Porreca)

113. In my opinion, the '202 Patent Pub. (Porreca) anticipates Claims 1-3, 5-11, 13-14, and 16 of the '241 Patent. Below is a table matching the claim terms in the '241 Patent to the corresponding disclosure in the '202 Patent Pub. (Porreca).

<u>'241 patent</u>	<u>'202 Patent Pub. (Porreca)</u>
“identifier site”	“differentiator tag sequence”
“indexing site”	“barcode”

1. Independent Claim 1 of the '241 Patent

(a) Limitation 1.p

114. Limitation 1.p recites a “method for detecting duplicate sequencing reads, the method comprising.” In my opinion, the '202 Patent Pub. (Porreca) discloses this limitation. For example, the '202 Patent Pub. (Porreca) teaches that differentiator tags may be used to identify multiple copies of the same nucleic acid sequence, *i.e.*, duplicate sequences. This is accomplished by uniquely tagging individual nucleic acid molecules. “In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve *the use of differentiator tag sequences to uniquely tag individual*

nucleic acid molecules.”¹⁸² The ’202 Patent Pub. (Porreca) also explains that “*each nucleic acid molecule may be tagged with a unique differentiator tag.*”¹⁸³ The use of uniquely tag individual nucleic acid molecules provides a method for detecting duplicate sequencing reads. “Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are *useful to determine whether target nucleic acid sequences identified in an assay are* from independently isolated target nucleic acids or *from multiple copies of the same target nucleic acid molecule* (e.g., due to bias in a preparative step, *for example, amplification*).”¹⁸⁴

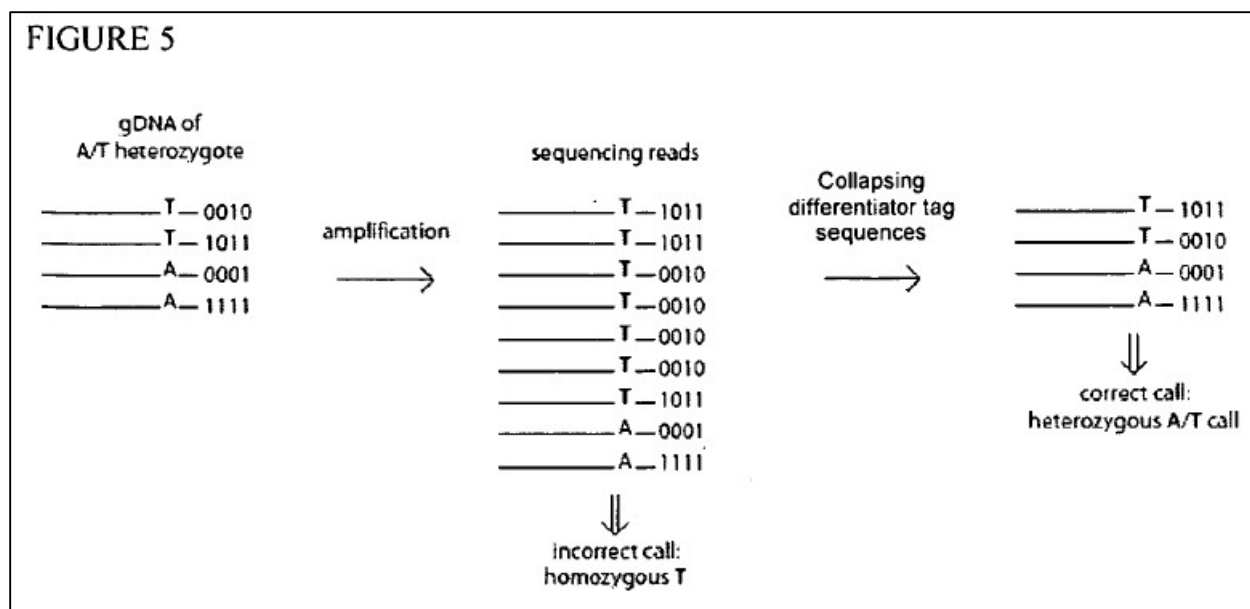
115. The ’202 Patent Pub. (Porreca)’s differentiator tags are attached to the nucleic acid fragments prior to amplification, as illustrated *infra* in Figure 5.

¹⁸² See Ex. 1006, ’202 Patent Pub. (Porreca) at [0023] (emphasis added).

¹⁸³ *Id.* at [0142] (emphasis added).

¹⁸⁴ *Id.* at [0024] (emphases added).

FIGURE 5



And because these unique tags are appended to nucleic acid fragments prior to PCR amplification, each identical sequence bearing the same differentiator tag will be detected as duplicate sequencing reads.

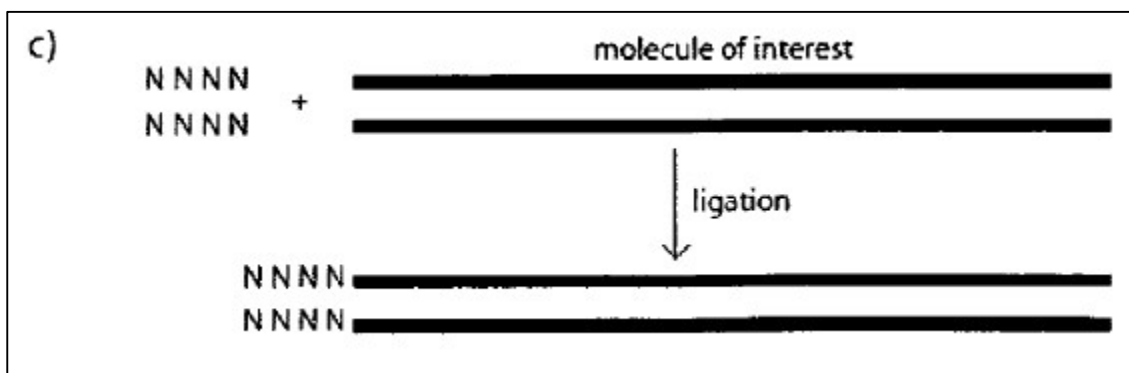
(b) Limitation 1.a

116. Limitation 1.a recites “obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor.” In my opinion, the ’202 Patent Pub. (Porreca) discloses this limitation. The ’202 Patent Pub. (Porreca) teaches that an adaptor containing a differentiator tag sequence may be appended to a nucleic acid sequence of interest, either through a ligation reaction or by a primer extension reaction.

The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, *nucleic acids comprising differentiator tag sequences may be incorporated by ligation*. This is a flexible method, because molecules having differentiator tag

sequence can be ligated to any blunt-ended nucleic acids. The sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed. ***Another method is to incorporate the differentiator tag sequence into a PCR primer***, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence ***followed by the PCR priming sequence (in 5' to 3' order)***. A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags.¹⁸⁵

117. Figure 4C of the '202 Patent Pub. (Porreca) illustrates an example where an adaptor comprising a differentiator tag sequence (NNNN) is ligated to the nucleic acid molecule of interest, as reproduced *infra*.



The '202 Patent Pub. (Porreca) also explains that “***FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.***”¹⁸⁶ Once the adaptor comprising a differentiator tag is appended

¹⁸⁵ *Id.* at [0150] (emphases added).

¹⁸⁶ *Id.* at [0140] (emphases added).

to the molecule of interest, the resulting product is amplified by PCR to produce amplicons. Figure 5 of the '202 Patent Pub. (Porreca) shows an example where nucleic acids with appended adaptors are amplified to generate amplicons, as reproduced *infra*.

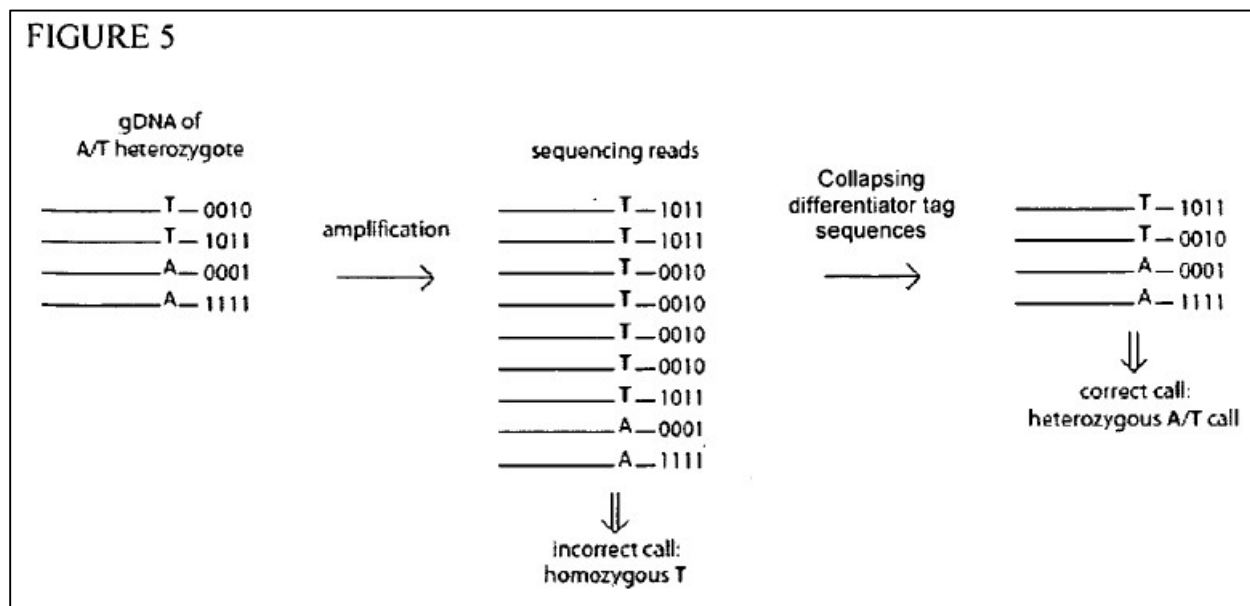


Figure 5 of the '202 Patent Pub. (Porreca) shows four nucleic acid molecules, each having an appended adaptor including a differentiator tag sequence (*e.g.*, “0010” or “1011”). These adaptor-appended nucleic acid molecules undergo “amplification” to produce several amplicons. This is shown under “sequencing reads” in the middle column of Figure 5 of the '202 Patent Pub. (Porreca).

(c) Limitation 1.b

118. Limitation 1.b recites “wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment.” In my opinion, the '202 Patent Pub. (Porreca) discloses this limitation. For example, the

'202 Patent Pub. (Porreca)'s differentiator tag satisfies the claimed "identifier site." The '241 Patent explains that an "identifier site" is "a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads."¹⁸⁷ The '202 Patent Pub. (Porreca) explains that these unique differentiator tags "may be created by using degenerate, random-sequence polynucleotides of defined length. . . For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N , where N is the length of the differentiator tag sequence."¹⁸⁸ The '202 Patent Pub. (Porreca) also explains that the "[d]ifferentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length."¹⁸⁹ For example, a differentiator tag sequence having a length 25 nucleotides provides 4^{25} , or 1.125×10^{15} unique identifiers. In other words, the '202 Patent Pub. (Porreca) discloses a sufficient number of identifier sequences to uniquely tag up to 1.125×10^{15} nucleic acid fragments.

119. The '202 Patent Pub. (Porreca) also discloses that each nucleic acid molecule is uniquely tagged with its own differentiator tag sequence. For example,

¹⁸⁷ See Ex. 1001, '241 Patent at 4:45-47.

¹⁸⁸ See Ex. 1006, '202 Patent Pub. (Porreca) at [0023] (emphasis added).

¹⁸⁹ *Id.* at [0141].

'202 Patent Pub. (Porreca) explains “*the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules.*”¹⁹⁰ The '202 Patent Pub. (Porreca) also explains that “[d]uring a library preparation process, *each nucleic acid molecule may be tagged with a unique differentiator tag sequence*”¹⁹¹ The '202 Patent Pub. (Porreca) further explains that “[i]n some embodiments, a *threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed* to determine the genotype of a subject.”¹⁹²

(d) Limitation 1.c

120. Limitation 1.c recites “sequencing the amplicons to generate sequence reads that include identifier and target sequences.” In my opinion, the '202 Patent Pub. (Porreca) discloses this limitation. For example, the '202 Patent Pub. (Porreca) teaches that amplicons comprising a differentiator tag appended to a target nucleic acid molecule (as described *supra* at Limitations 1.a and 1.b) are sequenced, and that the resulting sequencing reads contain both differentiator and target sequences.

During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that *permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of*

¹⁹⁰ *Id.* at [0023] (emphasis added).

¹⁹¹ *Id.* at [0142] (emphasis added); *see also id.* at [0143] (“A large library of unique differentiator tag sequences may be created”).

¹⁹² *Id.* at [0139].

interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence).¹⁹³

The '202 Patent Pub. (Porreca) also explains that the sequencing reads contain both differentiator and target sequences. “[T]he differentiator tag sequence permits the detection of bias based on the frequency with which *pairs of differentiator tag and target sequences are observed in a sequencing reaction.*”¹⁹⁴ The '202 Patent Pub. (Porreca) further explains that “[t]he sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence.”¹⁹⁵

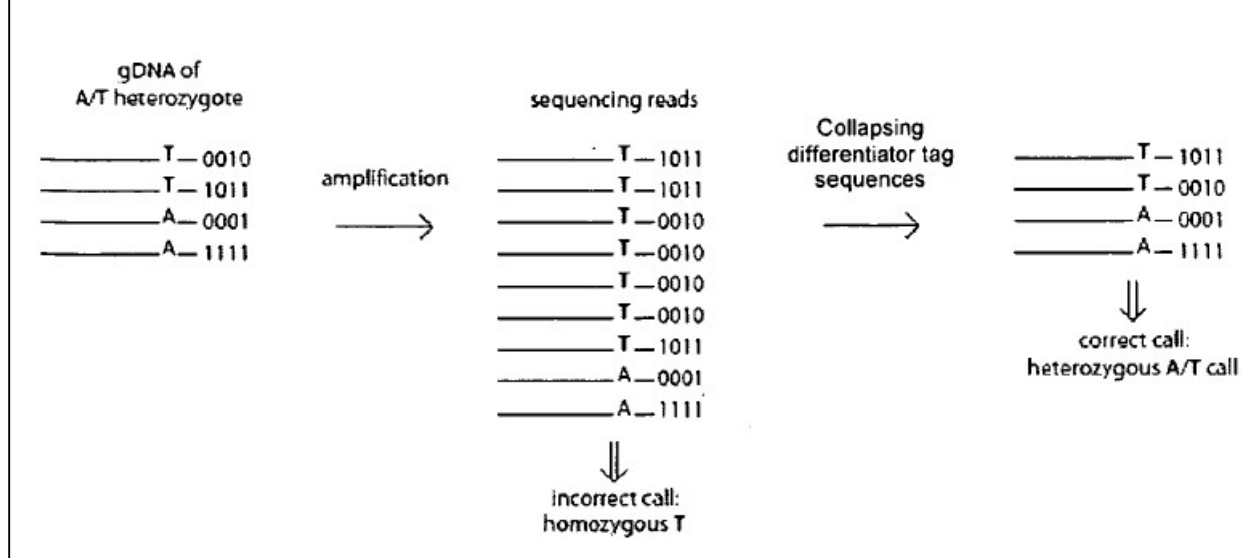
121. Figure 5 of the '202 Patent Pub. (Porreca) shows that both the differentiator tag and target nucleic acid sequences are sequenced together, as reproduced *infra*.

¹⁹³ *Id.* at [0142] (emphasis added).

¹⁹⁴ *Id.* at [0138] (emphasis added).

¹⁹⁵ *Id.* at [0138] (emphasis added).

FIGURE 5



The '202 Patent Pub. (Porreca) explains that “FIG. 5 depicts a non-limiting method for *genotyping based on target and differentiator tag sequences*. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls.”¹⁹⁶

(e) Limitation 1.d

122. Limitation 1.d recites “identifying sequence reads with identical identifier and target sequences as duplicates.” In my opinion, the '202 Patent Pub. (Porreca) discloses this limitation. For example, the '202 Patent Pub. (Porreca) teaches that differentiator tags can be used to identify whether particular nucleic acid molecules are copies of the same molecule that result from the PCR amplification process (*i.e.*, duplicates). Because each nucleic acid molecule is

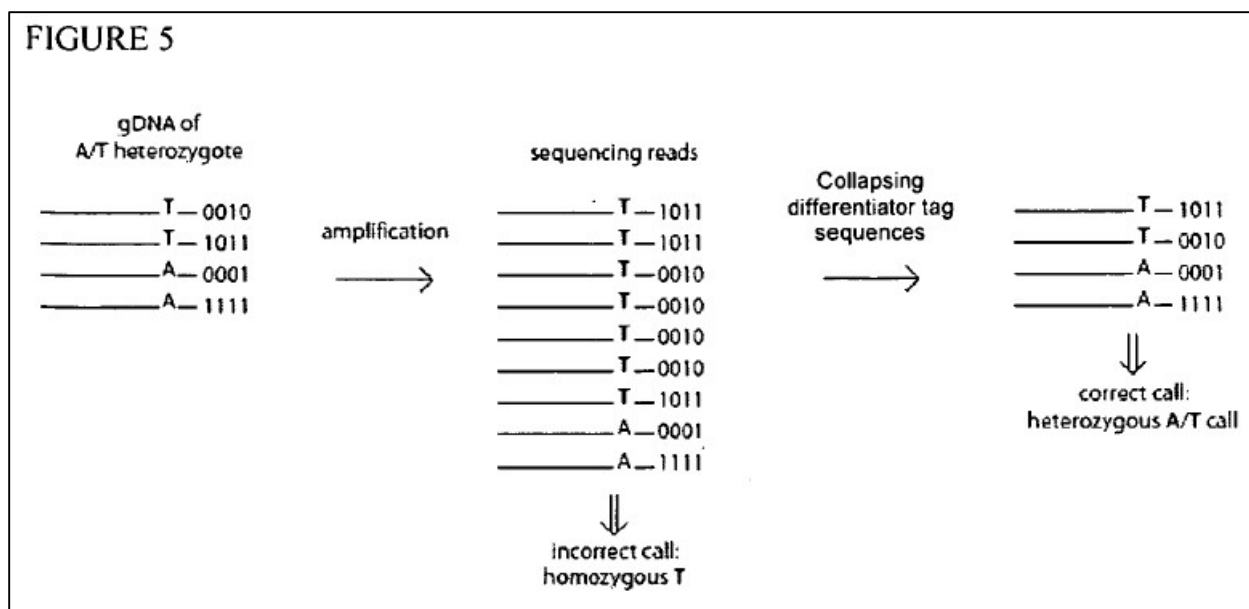
¹⁹⁶ *Id.* at [0148] (emphasis added).

tagged with a unique differentiator tag prior to amplification, molecules having identical nucleic acid sequences and bearing the same differentiator tag are duplicates of one another.

123. The '202 Patent Pub. (Porreca) explains that “sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids *or from multiple copies of the same target nucleic acid molecule* (e.g., due to bias in a preparative step, for example, *amplification*).”¹⁹⁷ One of ordinary skill in the art would have understood that in order to have “multiple copies of the same target nucleic acid molecule,” the target nucleic acid sequence must be identical. Figure 5 of the '202 Patent Pub. (Porreca) demonstrates that sequences having the same target nucleic acid sequence and differentiator tag are identified and collapsed, as reproduced *infra*.

¹⁹⁷ *Id.* at [0024] (emphases added).

FIGURE 5



124. The '202 Patent Pub. (Porreca) explains that “FIG. 5 depicts a non-limiting method for genotyping based on target and differentiator tag sequences. *Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls.*”¹⁹⁸ Figure 5 of the '202 Patent Pub. (Porreca) shows that duplicate sequences, such as two of the “T-1011” and three of the “T-0010” sequences, were identified as duplicates and collapsed into a single read. One of ordinary skill in the art would have known that the “T” represents a single nucleotide of the target sequence shown in Figure 5 of the '202 Patent Pub. (Porreca). One of ordinary skill in the art would have further known that, while not shown, the rest of the target sequence in the collapsed reads is also identical, as collapsing a read means removing sequences having the same target and

¹⁹⁸ *Id.* at [0148] (emphases added).

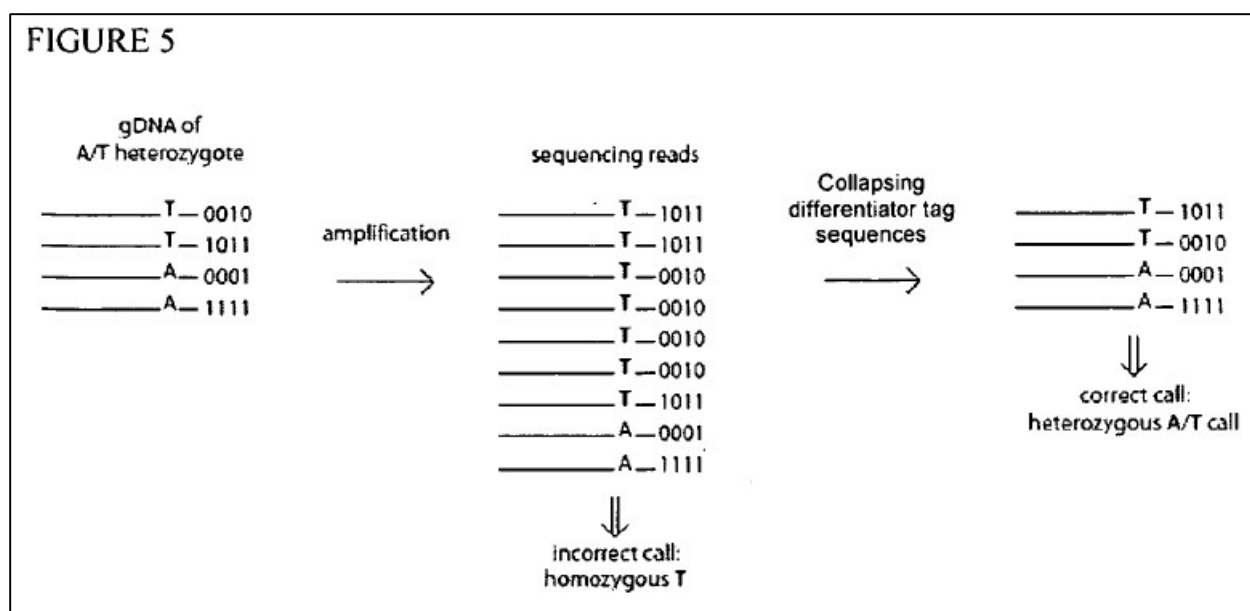
differentiator tag sequences. Indeed, this is what the '202 Patent Pub. (Porreca) instructs to do for their invention. “For any target:differentiator tag combination observed more than once, *all such reads are ‘collapsed’ into a single read before consensus basecalls are determined.*”¹⁹⁹

2. Claim 2 of the '241 Patent

125. Claim 2 recites “[t]he method of claim 1, further comprising removing a duplicate read from the sequence reads.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. As described *supra* at Limitation 1.d, the '202 Patent Pub. (Porreca) discloses the identification and “collapsing” of duplicate reads. One of ordinary skill in the art would have understood that “collapsing” the sequences means that the sequences having identical target and differentiator tag sequences are removed from the remainder of sequence reads. The removal of duplicate sequence reads is shown *infra* in Figure 5 of the '202 Patent Pub. (Porreca).

¹⁹⁹ *Id.* at [0298] (emphasis added).

FIGURE 5



126. There are four copies of the sequence read “T-0010,” all of which have the same target sequence (one nucleotide of which is “T”) and the same “0010” differentiator tag sequence. There are also three copies of the sequence read “T-1011,” all of which have the same target sequence (one nucleotide of which is “T”) and the same “1011” differentiator tag sequence. After the collapsing step, only a single sequence read copy of the “T-0010” product and “T-1011” product remain. Figure 5 of the ’202 Patent Pub. (Porreca) shows that the duplicate copies of sequence reads are removed from the analysis. As described *supra*, one of ordinary skill in the art would have understood that collapsing identical differentiator tag sequences in a single read would require discarding or removing the other sequences from that group. I note that the ’241 Patent also acknowledges that collapsing identical reads into a single read results in removing duplicate sequence reads. “[D]uplicate reads can be considered non-informative and *may be collapsed*

down to a single read, thus reducing the number of sequencing reads used in final analysis.”²⁰⁰ This understanding is consistent with other teachings of the ’202 Patent Pub. (Porreca) that sequences having identical target and differentiator sequences should be collapsed. “For any target:differentiator tag combination observed more than once, *all such reads are ‘collapsed’ into a single read before consensus basecalls are determined.*”²⁰¹ Thus, the ’202 Patent Pub. (Porreca)’s description of collapsing duplicate reads into a consensus read satisfies this claim.

3. Claim 3 of the ’241 Patent

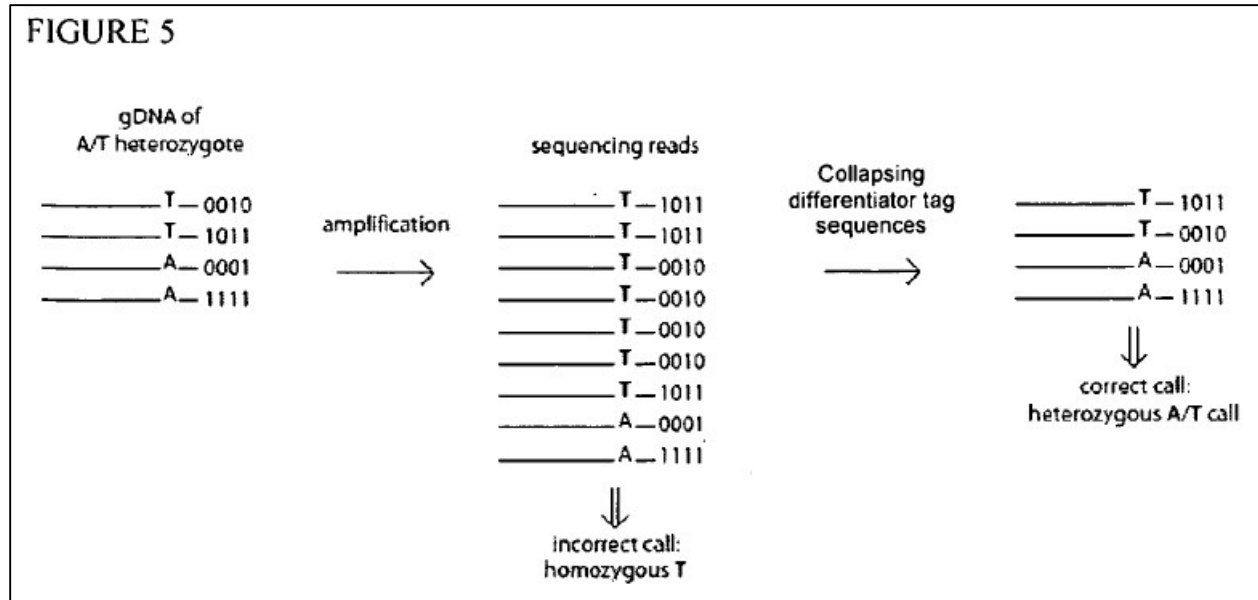
127. Claim 3 recites “[t]he method of claim 1, wherein the identifier site is sequenced with the target sequence.” In my opinion, the ’202 Patent Pub. (Porreca) discloses this claim. For example, the ’202 Patent Pub. (Porreca) teaches that the differentiator tag sequence and the target sequence can be sequenced together. “During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence *in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest . . .*”²⁰² The ’202 Patent Pub. (Porreca) also explains that “[i]n some embodiments, the *differentiator tag sequence permits the detection of bias*

²⁰⁰ See Ex. 1001, ’241 Patent at 1:42-45 (emphasis added).

²⁰¹ See Ex. 1006, ’202 Patent Pub. (Porreca) at [0298] (emphasis added).

²⁰² *Id.* at [0142] (emphasis added).

*based on the frequency with which pairs of differentiator tag and target sequences are observed in a sequencing reaction.*²⁰³ The '202 Patent Pub. (Porreca) additionally teaches that for these embodiments, “[t]he sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence.”²⁰⁴ Figure 5 of the '202 Patent Pub. (Porreca) further shows an embodiment where the differentiator tag is sequenced with the target sequence, as reproduced *infra*.



²⁰³ *Id.* at [0138] (emphasis added).

²⁰⁴ *Id.* at [0150] (emphasis added).

Thus, the '202 Patent Pub. (Porreca)'s description that the identifier site is sequenced with the target sequence satisfies the claim.

4. Claim 5 of the '241 Patent

128. Claim 5 recites “[t]he method of claim 1, wherein the amplicons include amplified fragments of cDNA.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. For example, the '202 Patent Pub. (Porreca) explains that its methods can be applied to a wide variety of nucleic acids, including cDNA.

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. . . . ***The nucleic acid may be*** DNA or RNA, such as genomic DNA, mitochondrial DNA, mRNA, ***cDNA***, rRNA, miRNA, or a combination thereof.”²⁰⁵

The '202 Patent Pub. (Porreca) further teaches that the differentiator tag may be incorporated at the cDNA level. “Another method to incorporate the differentiator tag sequences is as a tag on a gene-specific or ***poly-dT reverse-transcription primer***. ***This allows the differentiator tag sequence to be incorporated directly at the cDNA level.***”²⁰⁶ Thus, the '202 Patent Pub. (Porreca)'s description of cDNA or reverse transcribing RNA into cDNA satisfies the claim.

²⁰⁵ *Id.* at [0200] (emphases added).

²⁰⁶ *Id.* at [0150] (emphasis added).

5. Claim 6 of the '241 Patent

129. Claim 6 recites “[t]he method of claim 1, wherein the amplicons include amplified fragments of genomic DNA.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. For example, the '202 Patent Pub. (Porreca) teaches that genomic DNA may be used in its methods.

It should be appreciated that probes, primers, and other nucleic acids designed or used herein may be synthetic, natural, or a combination thereof. . . . Nucleic acids can be obtained from natural sources...*The nucleic acid may be* DNA or RNA, such as *genomic DNA*, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof.²⁰⁷

The '202 Patent Pub. (Porreca) also discloses embodiments analyzing genomic DNA.

This ensures that products of the preparative methods include only unique target:differentiator tag sequence; e.g., in a MIP reaction the capture step would undersample the total population of unique differentiator tag sequences in the MIP library. For example, an experiment utilizing 1 µg of *genomic DNA* will contain about ~150,000 copies of a diploid genome.”²⁰⁸

Thus, the '202 Patent Pub. (Porreca)'s description of a genomic DNA satisfies the claim.

²⁰⁷ *Id.* at [0200] (emphases added).

²⁰⁸ *Id.* at [0147] (emphasis added).

6. Claim 7 of the '241 Patent

130. Claim 7 recites “[t]he method of claim 1, wherein each adaptor includes a primer binding site.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. For example, the '202 Patent Pub. (Porreca) teaches that adaptors can include primer binding sites. “[A]dapters may be ligated onto the ends of the molecules of interest. *Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites.*”²⁰⁹ The '202 Patent Pub. (Porreca) also explains that “[t]he sequencing primers must be *incorporated subsequently* such that they sequence both the differentiator tag sequence and the target sequence. Alternatively, *the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed.*”²¹⁰

131. The '202 Patent Pub. (Porreca) also teaches that a “common adaptor sequence” may be used. One of ordinary skill in the art would have known that a common adaptor sequence, also known as a universal adaptor, would have a primer binding site to enable amplification and/or sequencing of all amplicons with a common primer sequence. The '202 Patent Pub. (Porreca) further explains the

²⁰⁹ *Id.* at [0149] (emphasis added).

²¹⁰ *Id.* at [0150] (emphases added).

orientation of the differentiator tag sequence associated with the common adaptor sequence.

*Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags.*²¹¹

Thus, the '202 Patent Pub. (Porreca)'s description of an adaptor that includes a primer binding site satisfies the claim.

7. Claim 8 of the '241 Patent

132. Claim 8 recites “[t]he method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. A universal target sequencing primer binding site is typically associated with a common adapter sequence to enable sequencing of all amplicons with a common sequencing primer. The '202 Patent Pub. (Porreca) explains that “[a]dapters often contain PCR primer sites (for amplification or emulsion PCR) and/or *sequencing primer sites*.”²¹² The '202 Patent Pub. (Porreca) also explains that “[t]he *sequencing primers must be incorporated subsequently* such that they sequence both the differentiator tag

²¹¹ *Id.* (emphases added).

²¹² *Id.* at [0149] (emphasis added).

sequence and the target sequence. Alternatively, *the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed.*"²¹³

133. The '202 Patent Pub. (Porreca) also discloses "common adaptor."²¹⁴ As described *supra*, one of ordinary skill in the art would have understood a common adaptor, also known as a universal adaptor, would have an "universal target sequencing primer binding site" to enable sequencing of all amplicons with a common primer sequence. The '202 Patent Pub. (Porreca) further explains the orientation of the differentiator tag sequence associated with the common adapter sequence.

*Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5' to 3' order). A differentiator tag sequence and adaptor sequence (which may contain the sequencing primer site) are incorporated as tags.*²¹⁵

²¹³ *Id.* at [0150] (emphases added).

²¹⁴ *See* Ex. 1001, '241 Patent at 9:63-67 (emphasis added). ("In some embodiments, the adaptor can be ligated to each of the nucleic acid fragments such that each of the nucleic acid fragments comprises the same adaptor. In other words, each of the nucleic acid fragments comprises a *common adaptor*.") (emphasis added).

²¹⁵ *See* Ex. 1006, '202 Patent Pub. (Porreca) at [0150] (emphases added).

Thus, the '202 Patent Pub. (Porreca)'s description of a primer binding site that is a universal target sequencing primer binding site satisfies the claim.

8. Claim 9 of the '241 Patent

134. Claim 9 recites “[t]he method of claim 1, wherein each unique identifier is at least 6 nucleotides in length.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. For example, the '202 Patent Pub. (Porreca) discloses differentiator tag sequences having various nucleotide lengths, including those with at least 6 nucleotides. “*Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up to 22, up to 23, up to 24, up to 25, or more nucleotides in length.*”²¹⁶ Thus, the '202 Patent Pub. (Porreca)'s description of the length of the unique identifier can be as long as 25 nucleotides satisfies the claim.

9. Claim 10 of the '241 Patent

135. Claim 10 recites “[t]he method of claim 1, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. For example, the '202 Patent Pub. (Porreca) teaches that its differentiator tags may be ligated to the target sequence. “The actual incorporation of the random differentiator tag

²¹⁶ *Id.* at [0141] (emphasis added).

sequences can be performed through various methods known in the art. For example, *nucleic acids comprising differentiator tag sequences may be incorporated by ligation.*²¹⁷ The '202 Patent Pub. (Porreca) explains “[a]lternatively, the *sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3’ end (as degenerate bases), so that only one ligation must be performed.*”²¹⁸

136. The '202 Patent Pub. (Porreca) teaches that the differentiator sequence may be appended to the target sequence using a priming extension reaction as well. “Another method is to incorporate the differentiator tag sequence into a PCR primer, such that the primer structure is arranged with the common adaptor sequence followed by the random differentiator tag sequence followed by the PCR priming sequence (in 5’ to 3’ order).”²¹⁹ Thus, the '202 Patent Pub. (Porreca)’s description of a ligation reaction and a primer extension reaction satisfies the claim.

10. Claim 11 of the '241 Patent

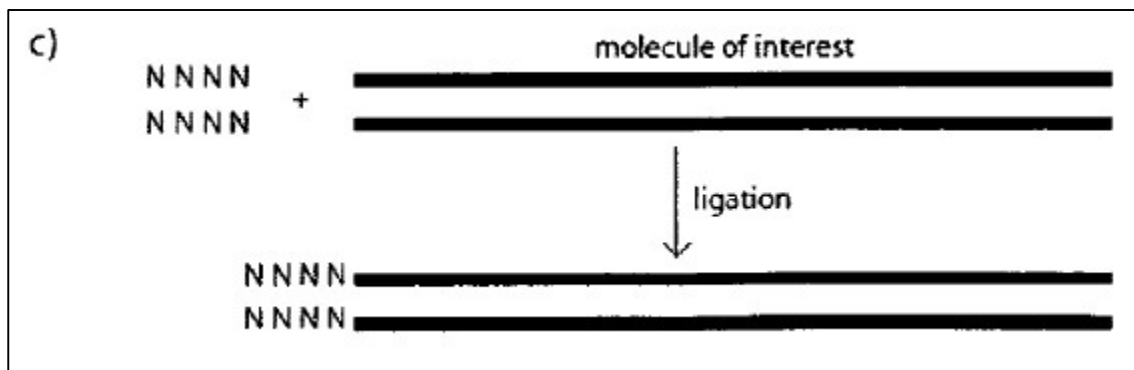
137. Claim 11 recites “[t]he method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. For example, the '202

²¹⁷ *Id.* at [0150] (emphasis added).

²¹⁸ *Id.* (emphasis added); *see also id.* at Figure 4c.

²¹⁹ *Id.* (emphasis added).

Patent Pub. (Porreca) teaches that “the sequencing adaptors can be synthesized with *the random differentiator tag sequences at their 3’ end (as degenerate bases), so that only one ligation must be performed.*”²²⁰ One of ordinary skill in the art would have understood that adaptors having a unique identifier at their 3’ end upon ligation to the nucleic acid fragments would result in a configuration whereby the unique identifier is at a junction between the adapter and the nucleic acid fragment. This configuration is also shown in Figure 4c of the ’202 Patent Pub. (Porreca), where the differentiator tag (NNNN) is at a junction between the adaptor and the molecule of interest, as reproduced *infra*.



Thus, the ’202 Patent Pub. (Porreca)’s description of each unique identifier is at a junction between the adaptor and the nucleic fragment satisfies the claim.

11. Claim 13 of the ’241 Patent

138. Claim 13 recites “[t]he method of claim 1, wherein each adaptor further comprises an indexing site.” In my opinion, the ’202 Patent Pub. (Porreca)

²²⁰ *Id.* (emphasis added).

discloses this claim. For example, the '202 Patent Pub. (Porreca) teaches adaptors may include “barcodes” to uniquely identify particular samples.

Adapters often contain PCR primer sites (for amplification or emulsion PCR) and/or sequencing primer sites. In addition, **barcodes** may be included, for example, to **uniquely identify individual samples** (e.g., patient samples) that may be mixed together. (See, e.g., US PTO Publication Number US 2007 /0020640A1 (McCloskey et al.).²²¹

One of ordinary skill in the art would have known that the '202 Patent Pub. (Porreca)'s “barcode” is appended to the target nucleic acid sequence and that all nucleic acids from the same sample would have an identical barcode. The '202 Patent Pub. (Porreca)'s barcode thus meets the claimed indexing site which is known as “multiplexing.”

139. Inclusion of this barcode allows for multiple samples to be run together, which one of ordinary skill in the art would have known is referred to as multiplex analysis, which the '202 Patent Pub. (Porreca) also discusses. “In some configurations, predetermined preparative steps and/or nucleic acid sequence analysis techniques are used **in multiplex analyses for a plurality of genetic loci in a plurality of samples**”²²² The '202 Patent Pub. (Porreca) also explains that “[i]n some embodiments, samples from **different subjects are tagged with subject-**

²²¹ *Id.* at [0149] (emphasis added).

²²² *Id.* at Abstract (emphasis added).

*specific (e.g., patient-specific) tags (e.g., unique sequence tags) so that the information from each product can be assigned to an identified subject.”*²²³

140. After amplification and sequencing, the sample can be “demultiplexed” by grouping together the sequencing reads that bear the same barcode. Moreover, the cited paragraph 149 *supra* in the ’202 Patent Pub. (Porreca) references the ’640 Patent Pub. (McCloskey), which also discloses the claimed indexing site. As described *supra* for Ground I at Limitation 1.a and Claims 13, the ’640 Patent Pub. (McCloskey)’s batch-stamp satisfies the claimed indexing site. Thus, the ’202 Patent Pub. (Porreca)’s description of a barcode sequence satisfies the claim.

12. Claim 14 of the ’241 Patent

141. Claim 14 recites “[t]he method of claim 1, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length.” In my opinion, the ’202 Patent Pub. (Porreca) discloses this claim. For example, the ’202 Patent Pub. (Porreca) discloses differentiator tag sequences having a variety of nucleotide lengths, including those with at least 6 nucleotides. “*Differentiator tag sequences may be, up to 5, up to 6, up to 7 up to 8, up to 9, up to 10*, up to 11, up to 12, up to 13, up to 14, up to 15, up to 16, up to 17, up to 18, up to 19, up to 20, up to 21, up

²²³ *Id.* at [0102] (emphasis added).

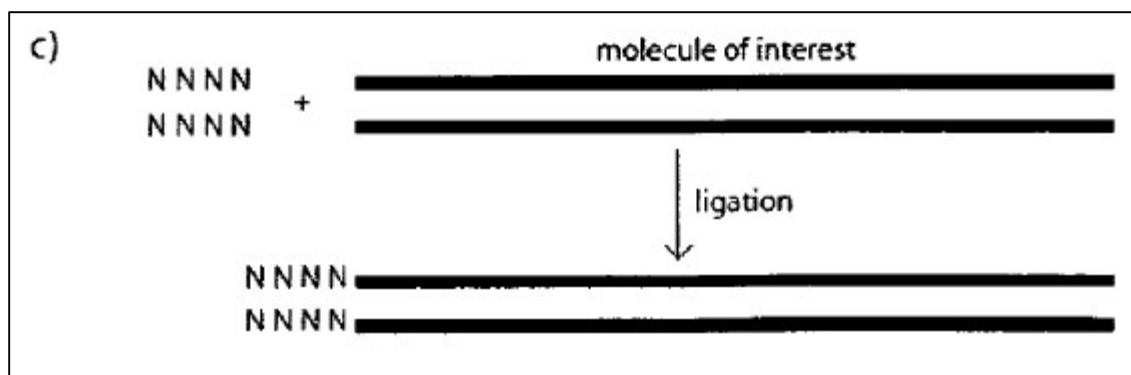
to 22, up to 23, up to 24, up to 25, or more nucleotides in length.”²²⁴ Thus, the ’202 Patent Pub. (Porreca)’s description that the unique identifier can be at least 5, 6, 7, 8, 9, or 10 nucleotides in length satisfies the claim.

13. Claim 15 of the ’241 Patent

142. Claim 15 recites “[t]he method of claim 1, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and performing an amplification reaction to obtain the amplicons.” In my opinion, the ’202 Patent Pub. (Porreca) discloses this claim. As described *supra* at Limitation 1.a, the ’202 Patent Pub. (Porreca) discloses the amplicons recited in Claim 1 of the ’241 Patent. As the ’202 Patent Pub. (Porreca) further discloses that these amplicons were generated from nucleic acid fragments, Claim 15 of the ’241 Patent is disclosed as well.

143. For example, the ’202 Patent Pub. (Porreca)’s Figure 4C shows that the differentiator tag is ligated to a “molecule of interest”:

²²⁴ *Id.* at [0141] (emphasis added).



One of ordinary skill would have understood this “molecule of interest” to be a nucleic acid fragment. The ’202 Patent Pub. (Porreca) further discloses that “FIG.4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.”²²⁵ One of ordinary skill would also have understood that “target sequences” refer to nucleic acid fragments. Indeed, one of ordinary skill would have understood that NGS sequencing reactions have a limit on the size of the analyzed nucleic acid,²²⁶ and thus would have understood that fragmenting is necessary prior appending the adaptor. The ’202

²²⁵ *Id.* at [0140].

²²⁶ See Ex. 1033, Quail *et al.*, *Improved protocols for the Illumina Genome Analyzer sequencing system*, CURRENT PROTOCOLS IN HUMAN GENETICS 18.2.1-18.2.27 (2009) (“Quail (2009)”) at 18.2.3. (“*Fragmentation. Cluster amplification is a relatively inefficient process, and there is an upper limit to the size fragments that will amplify on the flowcell surface.* Standard Illumina sequencing libraries currently tend to have *a fragment size of 200 to 500 bp*, excluding adapters.” (emphases added)).

Patent Pub. (Porreca) teaches fragmenting of nucleic acids through shearing prior to adaptor ligation and amplification in several examples.²²⁷

14. Claim 16 of the '241 Patent

144. This claim recites “[t]he method of claim 1, wherein the amplicons include amplified fragment of a nucleic acid from a first sample with an appended adaptor that includes a first index and amplified fragment of a nucleic acid from a second sample with an appended adaptor that includes a second index, and the method includes separating sequence reads by sample based on the first or second index.” In my opinion, the '202 Patent Pub. (Porreca) discloses this claim. As described *supra* at Claim 13 of the '241 Patent, the '202 Patent Pub. (Porreca)'s batch-stamp satisfies the claimed indexing site.

145. The '202 Patent Pub. (Porreca) further discloses the analysis of nucleic acids from different samples. For example, the '202 Patent Pub. (Porreca) teaches that “nucleic acids are analyzed *in biological samples obtained from a plurality of different subjects*.”²²⁸ The '202 Patent Pub. (Porreca) also teaches that “[i]n some embodiments, samples from *different subjects are tagged with subject-specific*

²²⁷ See Ex. 1006, '202 Patent Pub. (Porreca) at [0251]. (“***Mechanically shear***, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. ***DNA should be sheared into fragments several hundred basepairs in length.***”) (emphases added); see also *id.* at [0293], [0379].

²²⁸ *Id.* at [0057] (emphasis added).

(e.g., patient-specific) tags (e.g., unique sequence tags) so that the *information from each product can be assigned to an identified subject.*²²⁹

146. One of ordinary skill in the art would have known that nucleic acids bearing different subject-specific tags (or in other words, different barcodes) originated from different samples. One of ordinary skill in the art would have further known that the steps of this claim are necessarily performed in order to assign “the information from each product” “to an identified subject.”²³⁰ Thus, the ’202 Patent Pub. (Porreca)’s disclosure of tagging samples from different subjects and subsequently assigning information from each product to an identified subject satisfies this claim.

C. Ground III: Claims 1-3, 5-12, and 14-15 Are Anticipated by the ’389 PCT Appl. (Schmitt)

147. In my opinion, the ’389 PCT Appl. (Schmitt) anticipates Claims 1-3, 5-7, 9-12, and 14-15 of the ’241 Patent. Below is a table matching the claim terms in the ’241 Patent to the corresponding disclosure in the ’389 PCT Appl. (Schmitt).

<u>’241 patent</u>	<u>’389 PCT Appl. (Schmitt)</u>
“identifier site”	“single molecule identifier (SMI)”
“indexing site”	“barcode”

²²⁹ *Id.* at [0102] (emphases added).

²³⁰ *Id.*

1. Independent Claim 1 of the '241 Patent

(a) Limitation 1.p

148. Limitation 1.p recites a “method for detecting duplicate sequencing reads, the method comprising.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this limitation. For example, the '389 PCT Appl. (Schmitt) teaches the use of a “single molecule identifier (SMI)” and explains that it may be used to identify PCR duplicates. “[T]he SMI adaptor molecules may be used in methods for unambiguous identification of PCR duplicates. In order to restrict sequencing analysis to uniquely sequenced DNA fragments, *many sequencing studies include a step to filter out PCR duplicates.*”²³¹ For example, the '389 PCT Appl. (Schmitt) also explains that “[e]very duplicate that arises from a single strand of DNA will have the same SMI, and thus each strand in a DNA duplex pair generates a distinct, yet related population of PCR duplicates after amplification owing to the complementary nature of the SMIs on the two strands of the duplex.”²³² The '389 PCT Appl. (Schmitt)’s Example 1 also discloses identifying PCR duplicates. “[T]he SMI tags were used to group together PCR duplicates that arose from individual single-stranded DNA molecules and to create a consensus sequence

²³¹ See Ex. 1007, '389 PCT Appl. (Schmitt) at [0085] (emphases added).

²³² *Id.* at [0098] (emphasis added).

from the family of duplicates. . . . *The improved sequence coverage arose from use of the SMI to identify PCR duplicates.*”²³³

(b) Limitation 1.a

149. Limitation 1.a recites “obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor.” In my opinion, the ’389 PCT Appl. (Schmitt) discloses this limitation. For example, the ’389 PCT Appl. (Schmitt) discloses that an adaptor containing an SMI is appended to a target nucleic acid sequence and that this resulting product is PCR amplified, which produces the claimed amplicons.

*Such a method may include steps of ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule to form a double-stranded SMI-target nucleic acid complex; amplifying the double-stranded SMI-target nucleic acid complex, resulting in a set of amplified SMI-target nucleic acid products.*²³⁴

150. The ’389 PCT Appl. (Schmitt) further discloses appending an adaptor to a nucleic acid fragment.

[E]ach end of the double-stranded target nucleic acid molecule is ligated to an SMI adaptor molecule. The double-stranded target nucleic acid complex is then amplified by a method known in the art (e.g., a PCR or non-PCR method known in the art),

²³³ *Id.* at [0104] (emphases added).

²³⁴ *Id.* at [0010] (emphasis added).

resulting in a set of uniquely labeled, amplified SMI-target nucleic acid products.²³⁵

151. The '389 PCT Appl. (Schmitt)'s Example 1 additionally discloses the claimed amplicons.

Such a method may include steps of ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule to form a double-stranded SMI-target nucleic acid complex; *amplifying the double-stranded SMI-target nucleic acid complex, resulting in a set of amplified SMI-target nucleic acid products.*²³⁶

152. The '389 PCT Appl. (Schmitt) also discloses amplifying a fragment of a nucleic acid with an appended adaptor.

[E]ach end of the double-stranded target nucleic acid molecule is ligated to an SMI adaptor molecule. *The double-stranded target nucleic acid complex is then amplified by a method known in the art (e.g., a PCR or non-PCR method known in the art), resulting in a set of uniquely labeled, amplified SMI-target nucleic acid products.*²³⁷

The '389 PCT Appl. (Schmitt) further explains that “*375 ng adaptor-ligated DNA was PCR amplified with primers* AATGATACGGCGACCACCGAG (SEQ ID

²³⁵ *Id.* at [0074] (emphasis added).

²³⁶ *Id.* at [0010] (emphases added).

²³⁷ *Id.* at [0074] (emphasis added); *see also id.* at [0098]. (“*The individually tagged strands are then PCR amplified.*”) (emphasis added); *see also id.* at [0089]-[0095].

NO:3) and GTGACTGGAGTTCAGACGTGTGC (SEQ ID NO:4) using the Kappa high-fidelity PCR kit for 8 cycles with an annealing temperature of 60 C.”²³⁸

(c) Limitation 1.b

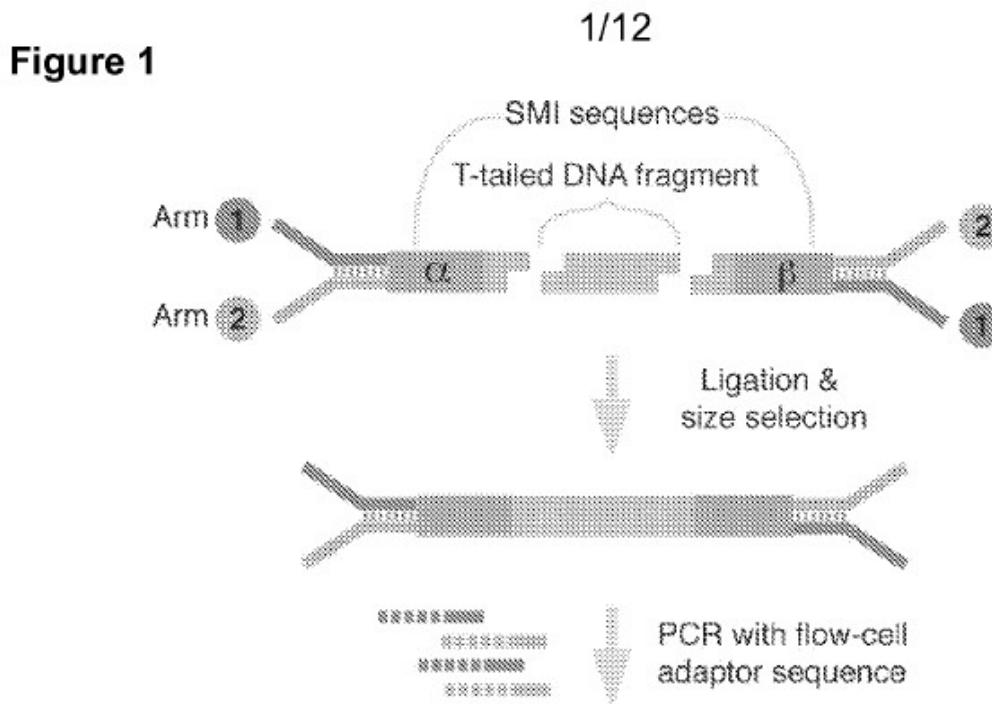
153. Limitation 1.b recites “wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment.” In my opinion, the ’389 PCT Appl. (Schmitt) discloses this limitation. For example, the ’389 PCT Appl. (Schmitt)’s SMI adaptor molecule satisfies the claimed “identifier site.” The ’389 PCT Appl. (Schmitt) discloses that the SMI sequence is made up of nucleotides and uniquely tags each DNA molecule.

[T]he SMI sequence (or “tag”) may be a double-stranded, complementary SMI sequence or a single-stranded SMI sequence. In some embodiments, the SMI adaptor molecule includes *an SMI sequence (or “tag”) of nucleotides that is degenerate or semi-degenerate. In some embodiments, the degenerate or semi-degenerate SMI sequence may be a random degenerate sequence.* A double-stranded SMI sequence includes *a first degenerate or semi-degenerate nucleotide n-mer sequence* and a second n-mer sequence that is complementary to the first degenerate or semi-degenerate nucleotide n-mer sequence, while a single-stranded SMI sequence includes *a first degenerate or semi-degenerate nucleotide n-mer sequence. The first and/or second degenerate or semi-degenerate nucleotide n-mer sequences may be any suitable length* to produce a sufficiently large number of unique tags to label a set of sheared DNA fragments from a segment of DNA. *Each n-mer sequence may be between approximately 3 to 20 nucleotides in length.* Therefore, each n-mer sequence may be approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides in length. In one embodiment, *the SMI*

²³⁸ *Id.* at [0093] (emphasis added).

sequence is a random degenerate nucleotide n-mer sequence which is 12 nucleotides in length. A 12 nucleotide SMI n-mer sequence that is ligated to each end of a target nucleic acid molecule, as described in the Example below, results in generation of up to 4^{24} (i.e., 2.8×10^{14}) distinct tag sequences.²³⁹

154. Figure 1 of the '389 PCT Appl. (Schmitt) shows an exemplary SMI adaptor that is ligated onto a target DNA fragment, as reproduced *infra*.



Regarding Figure 1, the '389 PCT Appl. (Schmitt) explains that “[b]ecause every adaptor contains *a unique, double-stranded, complementary n-mer random tag on each end (n-mer = 12 bp according to one embodiment)*, every DNA fragment becomes labeled with two distinct SMI sequences.”²⁴⁰ The '389 PCT Appl.

²³⁹ *Id.* at [0026] (emphases added).

²⁴⁰ *Id.* at [0012] (emphasis added).

(Schmitt) further explains that “[b]ecause the use of SMI tags (or ‘double-stranded SMI sequences’) allows every molecule to be uniquely labeled prior to PCR duplication, true PCR duplicates may be unambiguously identified by virtue of having a common (i.e., the same or identical) SMI sequence.”²⁴¹

155. The ’389 PCT Appl. (Schmitt)’s discussion of Example 1 also discloses this limitation.

In order to label or tag each of the strands of duplex DNA with *unique complementary tags*, adaptors which contain the standard sequences required for the Illumina HiSeq system were synthesized, but *with addition of a double-stranded, complementary SMI sequence (or “tag”) of 12 random nucleotides (or a random “degenerate sequence”) per strand*. Target DNA molecules having a random SMI sequence n-mer that is 12 nucleotides in length on each end will therefore have a *unique 24 nucleotide SMI sequence*.”²⁴²

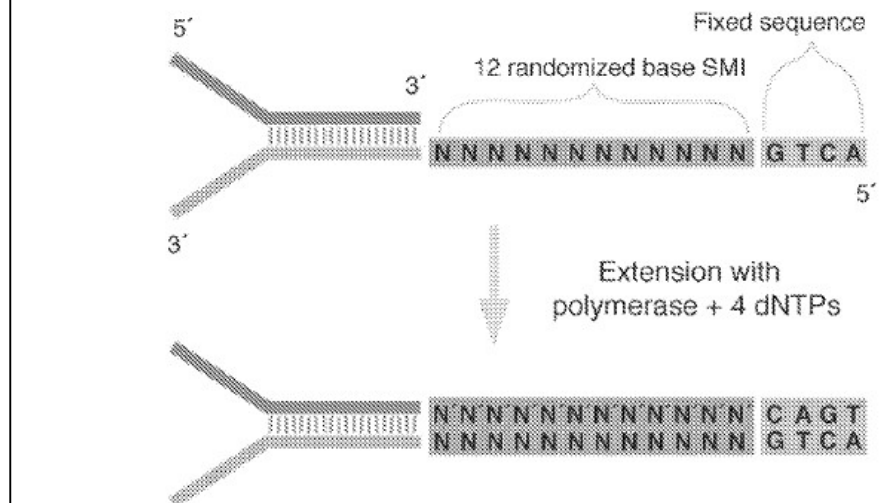
156. Figure 2 of the ’389 PCT Appl. (Schmitt) shows an exemplary 12 nucleotide SMI from Example 1, as reproduced *infra*.

²⁴¹ *Id.* at [0085] (emphasis added).

²⁴² *Id.* at [0101] (emphasis added); *see also id.* at [0088]-[0095], [0098]-[0099], [0104].

Figure 2

2/12



(d) Limitation 1.c

157. Limitation 1.c recites “sequencing the amplicons to generate sequence reads that include identifier and target sequences.” In my opinion, the ’389 PCT Appl. (Schmitt) discloses this limitation. For example, the ’389 PCT Appl. (Schmitt) teaches methods of appending an SMI adaptor to a target sequence and then amplifying and sequencing the resulting products.

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

amplified SMI-target nucleic acid products; and ***sequencing the amplified SMI-target nucleic acid products.***²⁴³

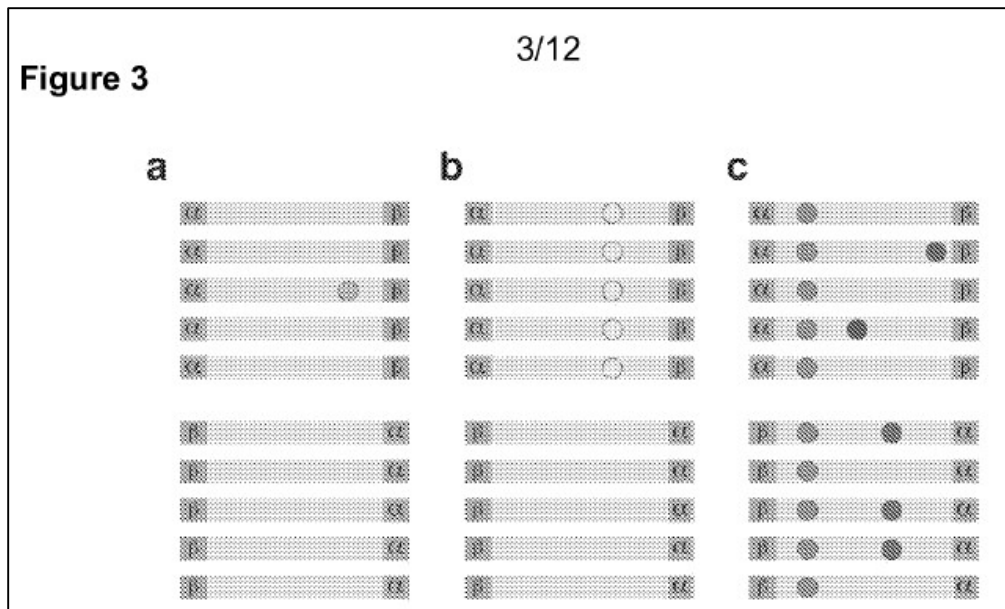
158. The '389 PCT Appl. (Schmitt) explains that the amplicons can be sequenced with any known methods in the art.

In one embodiment, each end of the double-stranded target nucleic acid molecule is ligated to an SMI adaptor molecule. The double-stranded target nucleic acid complex is then amplified by a method known in the art (e.g., a PCR or non-PCR method known in the art), ***resulting in a set of uniquely labeled, amplified SMI-target nucleic acid products. These products are then sequenced using any suitable method known in the art*** including, but not limited to, the Illumina sequencing platform, ABI SOLiD sequencing platform, Pacific Biosciences sequencing platform, 454 Life Sciences sequencing platform, Ion Torrent sequencing platform, Helicos sequencing platform, and nanopore sequencing technology.²⁴⁴

159. The '389 PCT Appl. (Schmitt) also discloses the sequencing of several SMI-nucleic acid products. For example, Figures 3a-c of the '389 PCT Appl. (Schmitt) show sequence reads that include both the unique identifier (α and β) and target sequence (the region in between α and β), as reproduced *infra*.

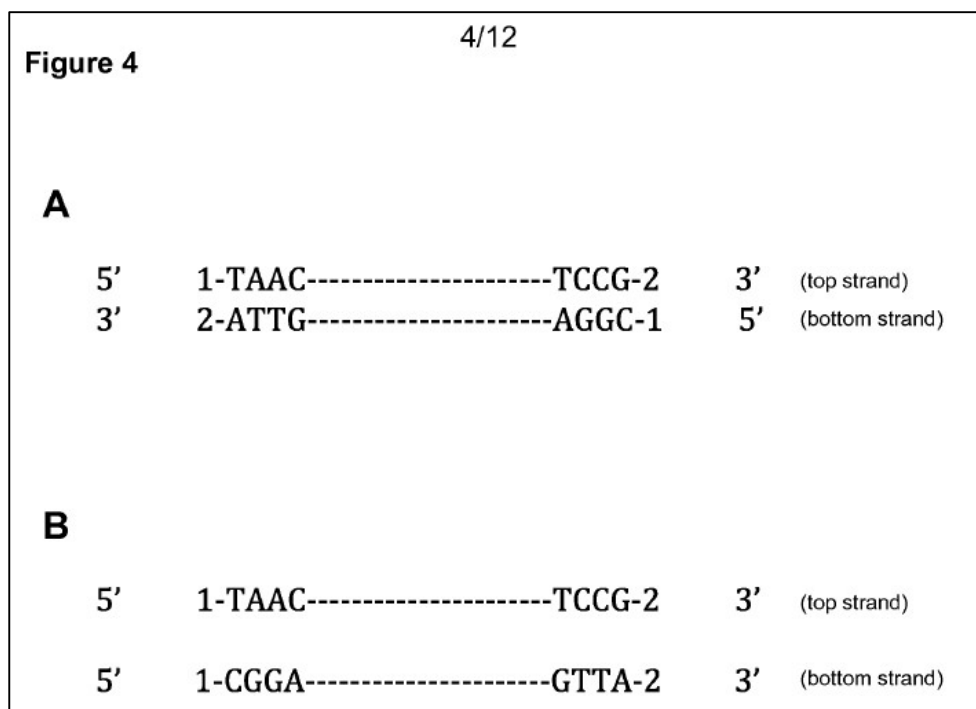
²⁴³ *Id.* at [0010] (emphasis added).

²⁴⁴ *Id.* at [0074] (emphasis added).



The '389 PCT Appl. (Schmitt) explains that “(a-c) shows *sequence reads (brown) sharing a unique set of SMI tags* are grouped into paired families with members having strand identifiers in either the $\alpha\beta$ or $\beta\alpha$ orientation.”²⁴⁵ Figure 4 of the '389 PCT Appl. (Schmitt) also shows an example of sequencing reads of both the identifier and target sequences, as reproduced *infra*.

²⁴⁵ *Id.* at [0014] (emphasis added).



The '389 PCT Appl. (Schmitt) explains Figure 4 as follows.

Figure 4 illustrates an example of how a SMI sequence with n-mers of 4 nucleotides in length (4-mers) are read by Duplex Consensus Sequencing (DCS) according to some embodiments. (A) shows the 4-mers with the PCR primer binding sites (or flow cell sequences) 1 and 2 indicated at each end. (B) shows the same molecules as in (A) but with the strands separated and the lower strand now written in the 5'-3' direction. *When these molecules are amplified with PCR and sequenced, they will yield the following sequence reads: The top strand will give a read 1 file of TAAC—and a read 2 file of GCCA---*. Combining the read 1 and read 2 tags will give TAACCGGA as the SMI for the top strand. The bottom strand will give a read 1 file of CGGA---- and a read 2 file of TAAC---. Combining the read 1 and read 2 tags will give CGGATAAC as the SMI for the bottom strand.”²⁴⁶

²⁴⁶ *Id.* at [0014] (emphasis added).

160. With respect to Example 1, the '389 PCT Appl. (Schmitt) explains that SMI-ligated DNA was PCR amplified and sequenced.²⁴⁷ “[M]embers of the same PCR ‘family’ are then grouped together by virtue of having a common (i.e., the same) SMI tag sequence. *The sequences of uniquely tagged PCR duplicates are subsequently compared in order to create a PCR consensus sequence.*”²⁴⁸ One of ordinary skill in the art would have understood that the “sequences of uniquely tagged PCR duplicates” refers to nucleic acid products having identical target nucleic acid sequences, as the nucleic acid products would not be referred to as “PCR duplicates” if they were not identical. Furthermore, Example 1’s protocol in the '389 PCT Appl. (Schmitt) “was *used to sequence DNA* isolated from normal colonic mucosa” and included sequence reads from tagged, amplicons with both SMI and target nucleic acid sequences.²⁴⁹

(e) Limitation 1.d

161. Limitation 1.d recites “identifying sequence reads with identical identifier and target sequences as duplicates.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this limitation. For example, the '389 PCT Appl. (Schmitt) teaches that identical target sequences that also bear the same SMI sequence are

²⁴⁷ Generally, *see id.* at [0098]-[0099].

²⁴⁸ *Id.* at [0099] (emphasis added); *see also id.* at [0096].

²⁴⁹ *Id.* at [0103]-[0105].

duplicates of one another. For example, the '389 PCT Appl. (Schmitt) teaches that “SMI adaptor molecules may be used in methods *for unambiguous identification of PCR duplicates*.”²⁵⁰ The '389 PCT Appl. (Schmitt) also explains that duplicate sequence read will have identical identifier and target sequences.

Every duplicate that arises from a single strand of DNA will have the same SMI, and thus each strand in a DNA duplex pair generates a distinct, yet related population of PCR duplicates after amplification *owing to the complementary nature of the SMIs on the two strands of the duplex*.²⁵¹

The '389 PCT Appl. (Schmitt) further explains that duplicate sequence read will have identical identifier and target sequences (*i.e.*, identical shear points).

However sequence reads with identical shear points can also reflect distinct molecules because there are a limited number of possible shear points at any given genomic location, and with increasing sequencing depth, recurrent shear points are increasingly likely to be seen. Because the use of SMI tags (or “double-stranded SMI sequences”) allows every molecule to be uniquely labeled prior to PCR duplication, *true PCR duplicates may be unambiguously identified by virtue of having a common (i.e., the same or identical) SMI sequence*.²⁵²

162. The '389 PCT Appl. (Schmitt) further discloses that duplicates were identified and then collapsed to generate a consensus read. “[M]embers of the same PCR ‘family’ are then grouped together by virtue of having a common (*i.e.*, the

²⁵⁰ *Id.* at [0085] (emphasis added).

²⁵¹ *Id.* at [0098] (emphases added).

²⁵² *Id.* at [0085] (emphases added).

same) SMI tag sequence. The sequences of uniquely tagged PCR duplicates are subsequently *compared in order to create a PCR consensus sequence.*²⁵³ The '389 PCT Appl. (Schmitt) further discloses that sequencing positions were discounted from the consensus read where “fewer than 90% of the sequences at that position in the consensus group *had the identical sequence.*”²⁵⁴ One of ordinary skill in the art would have known that the target sequences in the duplicates are identical as that is required to generate a consensus read. These disclosures thus further show that both the target sequence and SMI sequence are used to identify PCR duplicates in the '389 PCT Appl. (Schmitt).

2. Claim 2 of the '241 Patent

163. Claim 2 recites “[t]he method of claim 1, further comprising removing a duplicate read from the sequence reads.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this claim. For example, the '389 PCT Appl. (Schmitt)'s Example 1 demonstrates the removal of duplicate reads by collapsing the duplicate reads into a single consensus sequence read. “*Reads having common (i.e., identical) SMI sequences were grouped together, and were collapsed to generate*

²⁵³ *Id.* at [0099] (emphases added).

²⁵⁴ *Id.* at [0096] (emphasis added).

a consensus read.”²⁵⁵ The data processing workflow of Example 1 of the ’389 PCT Appl. (Schmitt) also includes the following steps.

6. *Group together reads which have identical 24 nt SMIs.*

7. *Collapse to SMI consensus reads*, scoring only positions with 3 or more SMI duplicates and >90% sequence identity among the duplicates.²⁵⁶

As described *supra*, one of ordinary skill in the art would have understood that collapsing duplicate reads into a consensus read removes those duplicate reads from the analysis. I note that the ’241 Patent also acknowledges that collapsing identical reads into a single read results in removing duplicate sequence reads. “[D]uplicate reads can be considered non-informative and *may be collapsed down to a single read, thus reducing the number of sequencing reads used in final analysis.*”²⁵⁷

Thus, the ’389 PCT Appl. (Schmitt)’s description of collapsing duplicate reads into a consensus read satisfies this claim.

3. Claim 3 of the ’241 Patent

164. Claim 3 recites “[t]he method of claim 1, wherein the identifier site is sequenced with the target sequence.” In my opinion, the ’389 PCT Appl. (Schmitt) discloses this claim. For example, the ’389 PCT Appl. (Schmitt) discloses the

²⁵⁵ *Id.* at [0096] (emphasis added).

²⁵⁶ *Id.* (emphases added).

²⁵⁷ *See* Ex. 1001, ’241 Patent at 1:42-45 (emphasis added).

following method, which involves the sequencing of the SMI sequence with the target sequence.

Such a method may include steps of ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule to form a double-stranded SMI-target nucleic acid complex; amplifying the double-stranded SMI-target nucleic acid complex, resulting in a set of amplified SMI-target nucleic acid products; and *sequencing the amplified SMI-target nucleic acid products*.²⁵⁸

165. The embodiments shown in Figures 3 and 4 of the '389 PCT Appl. (Schmitt) satisfy this limitation as well. For example, Figure 3a-c of the '389 PCT Appl. (Schmitt) "*shows sequence reads (brown) sharing a unique set of SMI tags are grouped into paired families* with members having strand identifiers in either the $\alpha\beta$ or $\beta\alpha$ orientation."²⁵⁹ The '389 PCT Appl. (Schmitt) also explains that when the molecules shown in Figures 4A and 4B "are amplified with PCR and sequenced, they will yield the following sequence reads: *The top strand will give a read 1 file of TAAC—and a read 2 file of GCCA---*."²⁶⁰ The '389 PCT Appl. (Schmitt) further explains that combining the reads "*will give TAACCGGA as the SMI for the top strand*" and "*CGGATAAC as the SMI for the bottom strand*."²⁶¹ The reads

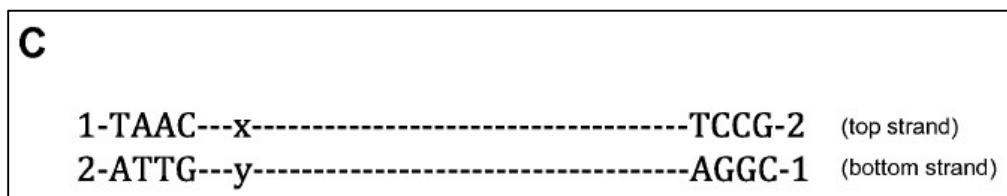
²⁵⁸ See Ex. 1007, '389 PCT Appl. (Schmitt) at [0010] (emphasis added).

²⁵⁹ *Id.* at [0014] (emphasis added).

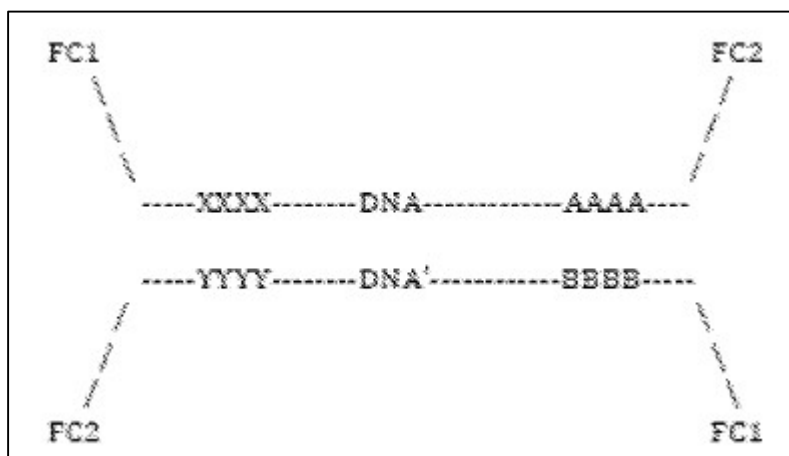
²⁶⁰ *Id.* at [0015] (emphasis added).

²⁶¹ *Id.* (emphasis added).

wherein the identifier site is sequenced with the target sequence are shown in Figure 4C of the '389 PCT Appl. (Schmitt), as reproduced *infra*.



The '389 PCT Appl. (Schmitt)'s illustration of Product II²⁶² shows another example of the identifier site sequenced with the target sequence), as reproduced *infra*.



This sequence read includes identifier (*i.e.*, XXXX and AAAA for the top strand and BBBB and YYYY for the bottom strand) and target sequences (*i.e.*, “DNA” for the top strand and “DNA'” for the bottom strand).²⁶³

166. The sequence read in Example 1 of the '389 PCT Appl. (Schmitt) additionally meets this limitation. The '389 PCT Appl. (Schmitt) discloses that

²⁶² *Id.* at [0042] (Product II).

²⁶³ *Id.* at [0045].

“[r]eads with intact SMI adaptors include a 12 nucleotide random sequence, followed by a 5 nucleotide fixed sequence” and that the *“SMI sequence from both the forward and reverse sequencing reads* (i.e., the first and second degenerate n-mer sequences) was computationally added to the read header, and the fixed sequence removed.”²⁶⁴ The ’389 PCT Appl. (Schmitt) further explains that “[r]eads having common (i.e., identical) SMI sequences were grouped together, and *were collapsed to generate a consensus read.*”²⁶⁵ As described *supra*, one of ordinary skill in the art would have known from these data processing steps that Example 1 of the ’389 PCT Appl. (Schmitt) to collapse sequence reads into in a single consensus read would require the sequencing reads with both identical identifier and target sequences. Thus, the ’389 PCT Appl. (Schmitt)’s description that the identifier site is sequenced with the target sequence satisfies the claim.

4. Claim 5 of the ’241 Patent

167. Claim 5 recites “[t]he method of claim 1, wherein the amplicons include amplified fragments of cDNA.” In my opinion, the ’389 PCT Appl. (Schmitt) discloses this claim. For example, the ’389 PCT Appl. (Schmitt) teaches that its methods are applicable to the analysis of RNA molecules.

[T]he SMI adaptor molecules and methods of use described herein are applicable to genotyping and other applications related

²⁶⁴ *Id.* at [0096] (emphases added).

²⁶⁵ *Id.* (emphasis added).

to sequencing of DNA molecules, *but are also applicable to RNA sequencing applications* such as for sequencing of double-stranded RNA viruses.²⁶⁶

The '389 PCT Appl. (Schmitt) further discloses that a ligated RNA molecule “*may be reverse transcribed into DNA*,” and then sequenced as a double-stranded target DNA molecule.”²⁶⁷ One of ordinary skill in the art would have known that reverse transcribing RNA results in cDNA, and that amplifying this product would have produced the claimed fragments of cDNA. Thus, the '389 PCT Appl. (Schmitt)'s description of reverse transcribing RNA into cDNA satisfies the claim.

5. Claim 6 of the '241 Patent

168. Claim 6 recites “[t]he method of claim 1, wherein the amplicons include amplified fragments of genomic DNA.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this claim. For example, the '389 PCT Appl. (Schmitt)'s Example 1 utilizes genomic DNA. “DNA isolation. *Genomic DNA was isolated* from normal human colonic mucosa by sodium iodide extraction.”²⁶⁸ The '389 PCT Appl. (Schmitt) also explains that “[g]enomic DNA was isolated from a derivative of *Saccharomyces cerevisiae* strain SC288 by standard methods.”²⁶⁹ Thus, the '389 PCT Appl. (Schmitt)'s description of a genomic DNA satisfies the claim.

²⁶⁶ *Id.* at [0072] (emphasis added).

²⁶⁷ *Id.* (emphasis added).

²⁶⁸ *Id.* at [0089] (emphasis added).

²⁶⁹ *Id.* at [0114] (emphasis added).

6. Claim 7 of the '241 Patent

169. Claim 7 recites “[t]he method of claim 1, wherein each adaptor includes a primer binding site.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this claim. For example, the '389 PCT Appl. (Schmitt) teaches that a primer binding site may be included in the adaptor. “Optionally, *the SMI adaptor molecule further includes at least two PCR primer binding sites, at least two sequencing primer binding sites, or both.*”²⁷⁰ The '389 PCT Appl. (Schmitt) also discloses the use of “Y-adaptors” that include a primer binding site.

In some embodiments, the SMI adaptor molecule is a “Y-shaped” adaptor, which allows both strands to be independently amplified by a PCR method prior to sequencing because both the *top and bottom strands have binding sites for PCR primers FC1 and FC2 as shown in the examples below.*”²⁷¹

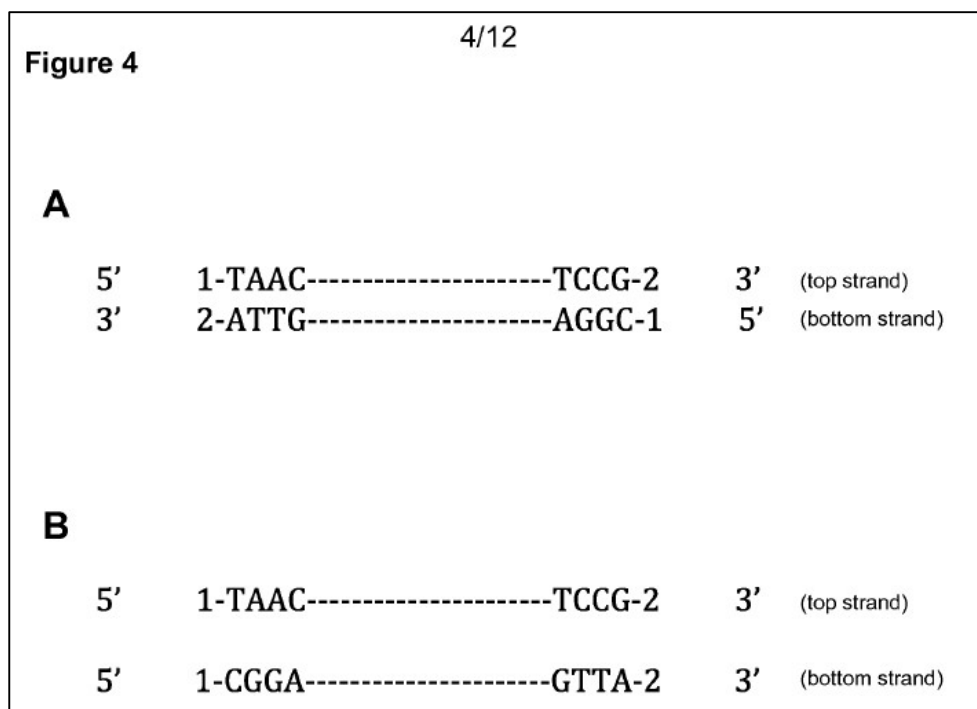
The '389 PCT Appl. (Schmitt) further teaches that “the *SMI adaptor molecule may include at least two PCR primer or ‘flow cell’ binding sites*: a forward PCR primer binding site (or a ‘flow cell 1’ (FC1) binding site); and a reverse PCR primer binding site (or a ‘flow cell 2’ (FC2) binding site).”²⁷²

170. The '389 PCT Appl. (Schmitt)’s Figures 4A and 4B also show primer binding sites, as reproduced *infra*.

²⁷⁰ *Id.* at [0022] (emphasis added).

²⁷¹ *Id.* at [0022] (emphasis added).

²⁷² *Id.* at [0031] (emphasis added).



The '389 PCT Appl. (Schmitt) explains that Figure 4A “*shows the 4-mers with the PCR primer binding sites (or flow cell sequences) 1 and 2 indicated at each end,*” and that Figure 4B “*shows the same molecules as in (A) but with the strands separated* and the lower strand now written in the 5'-3' direction.”²⁷³ The '389 PCT Appl. (Schmitt)'s Claim 5 additionally discloses this claim. “*The single molecule identifier adaptor molecule of claim 1, further comprising at least two PCR primer binding sites, or at least two sequencing primer binding sites, or both.*”²⁷⁴ Thus, the '389 PCT Appl. (Schmitt)'s description of an adaptor that includes a primer binding site satisfies the claim.

²⁷³ *Id.* at [0015] (emphases added).

²⁷⁴ *Id.* at Claim 5 (emphasis added).

7. Claim 8 of the '241 Patent

171. Claim 8 recites “[t]he method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site.” In my opinion, the ’389 PCT Appl. (Schmitt) discloses this claim. For example, the ’389 PCT Appl. (Schmitt) teaches that the SMI adaptor molecule can include a sequencing primer binding site, which meets the limitation *universal target sequencing primer binding site*.

The SMI adaptor molecule may also include at least two sequencing primer binding sites, each corresponding to a sequencing read. Alternatively, the sequencing primer binding sites may be added in a separate step by inclusion of the necessary sequences as tails to the PCR primers, or by ligation of the needed sequences.²⁷⁵

172. The ’389 PCT Appl. (Schmitt)’s discussion of Figure 4c also teaches a universal target sequencing primer binding site from which sequence read data are produced.

(C) illustrates the orientation of paired strand mutations in DCS. In the initial DNA duplex shown in Figures 4A and 4B, a mutation “x” (which is paired to a complementary nucleotide “y”) is shown on the left side of the DNA duplex. The “x” will appear in *read 1*, and the complementary mutation on the opposite strand, “y,” will appear in *read 2*. Specifically, this would appear as “x” in both *read 1 and read 2 data*, because “y” in read 2 is read out as “x” *by the sequencer owing to the nature*

²⁷⁵ *Id.* at [0031] (emphasis added).

of the sequencing primers, which generate the complementary sequence during read 2.²⁷⁶

173. The '389 PCT Appl. (Schmitt)'s claims also disclose the claimed sequencing primer binding site as well.

Claim 5. *The single molecule identifier adaptor molecule of claim 1, further comprising at least two PCR primer binding sites, or at least two sequencing primer binding sites, or both.*²⁷⁷

Claim 28. The method of claim 22, further comprising at least two PCR primer binding sites, *at least two sequencing primer binding sites*, a double-stranded fixed reference sequence or a combination thereof.²⁷⁸

As described *supra*, one of ordinary skill in the art would have understood a common adaptor, also known as a universal adaptor, would have an “universal target sequencing primer binding site” to enable sequencing of all amplicons with a common primer sequence. Thus, the '389 PCT Appl. (Schmitt)'s description of a primer binding site that is a universal target sequencing primer binding site satisfies the claim.

8. Claim 9 of the '241 Patent

174. Claim 9 recites “[t]he method of claim 1, wherein each unique identifier is at least 6 nucleotides in length.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this claim. For example, the '389 PCT Appl. (Schmitt)

²⁷⁶ *Id.* at [0015] (emphases added).

²⁷⁷ *Id.* at Claim 5 (emphases added).

²⁷⁸ *Id.* at Claim 28 (emphasis added).

discloses that a SMI adaptor may be a nucleotide n-mer sequence up to 20 nucleotides in length.

Each n-mer sequence may be between approximately 3 to 20 nucleotides in length. Therefore, each n-mer sequence may be approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides in length. In one embodiment, *the SMI sequence is a random degenerate nucleotide n-mer sequence which is 12 nucleotides in length.* A 12 nucleotide SMI n-mer sequence that is ligated to each end of a target nucleic acid molecule, as described in the Example below, results in generation of up to 4^{24} (i.e., 2.8×10^{14}) distinct tag sequences”²⁷⁹

Example 1 of the '389 PCT Appl. (Schmitt) also involves the use of “a 24-nucleotide SMI” that “*consists of two 12-nucleotide sequences* that can be designated XY.”²⁸⁰ Thus, the '389 PCT Appl. (Schmitt)’s description of the length of the unique identifier can be as long as 20 nucleotides satisfies the claim.

9. Claim 10 of the '241 Patent

175. Claim 10 recites “[t]he method of claim 1, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this claim. For example, the '389 PCT Appl. (Schmitt) discloses that the SMI adaptor can be ligated to the target nucleic acid sequence.

SMI adaptor molecule includes a single molecule identifier (SMI) sequence which comprises a degenerate or semi-degenerate DNA sequence; and *an SMI ligation adaptor that*

²⁷⁹ *Id.* at [0026] (emphases added).

²⁸⁰ *Id.* at [0100] (emphasis added).

*allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence.*²⁸¹

The '389 PCT Appl. (Schmitt) also teaches a sequencing method that “may include steps of *ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule* to form a double-stranded SMI-target nucleic acid complex.”²⁸² The '389 PCT Appl. (Schmitt) further explains that “[i]n some embodiments, *the SMI adaptor molecules are ligated to both ends of a target nucleic acid molecule.*”²⁸³

176. The SMI adaptor in the '389 PCT Appl. (Schmitt)'s Example 1 was also ligated to the target nucleic acid sequence. “*The custom adaptors were ligated by combining 750 ng of T-tailed DNA* with 250 pmol adaptors in a reaction containing 3000 units T4 DNA ligase, 50 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM DTT, 1 mM ATP.”²⁸⁴ Thus, the '389 PCT Appl. (Schmitt)'s description of a ligation reaction satisfies the claim.

10. Claim 11 of the '241 Patent

177. Claim 11 recites “[t]he method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.” In my opinion, the

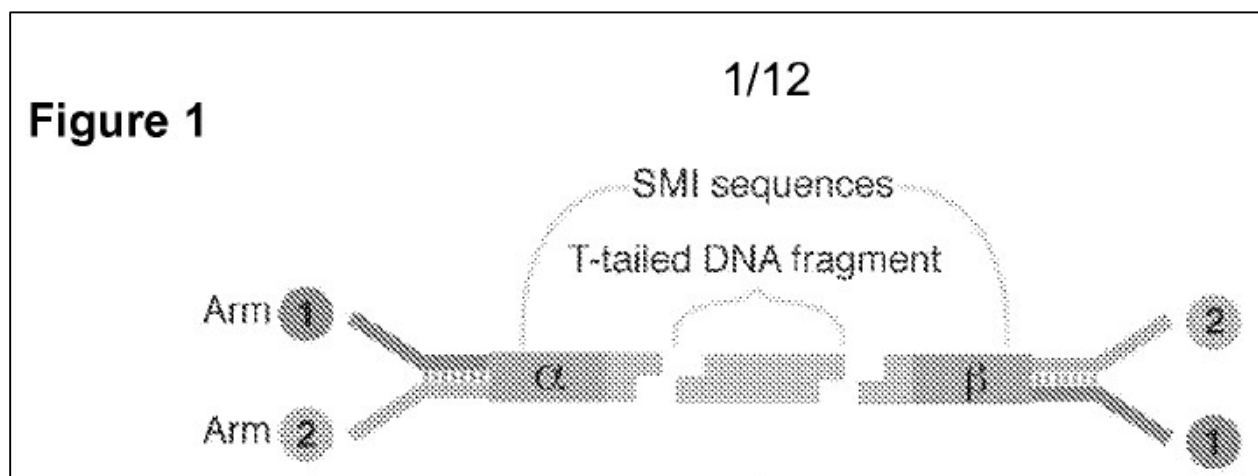
²⁸¹ *Id.* at [0009] (emphasis added).

²⁸² *Id.* at [0010] (emphasis added).

²⁸³ *Id.* at [0028] (emphasis added); *see also id.* [0072], [0074].

²⁸⁴ *Id.* at [0010] (emphasis added).

'389 PCT Appl. (Schmitt) discloses this limitation. I understand that during prosecution of a related patent, the examiner found that the term “junction” is considered as any region of the adaptor.²⁸⁵ The '389 PCT Appl. (Schmitt)'s Figure 1, which shows Example 1's duplex consensus sequencing method, illustrates that the SMI is located at the junction between the adapter and the nucleic acid fragment, as reproduced *infra*.



178. The '389 PCT Appl. (Schmitt) explains that the unique identifier is between the adaptor and the nucleic fragment.

Figure 1 illustrates an overview of Duplex Consensus Sequencing. Sheared double-stranded DNA that has been end-repaired and T-tailed is combined with A-tailed SMI adaptors and ligated according to one embodiment. . . . *[E]very DNA fragment becomes labeled with two distinct SMI sequences (arbitrarily designated α and β in the single capture event shown).* . . . Those derived from one strand will have *the α SMI*

²⁸⁵ See Ex. 1013,' 357 Patent File History, December 8, 2020 nonfinal rejection at 3. (“Applicant did not define the term “junction”, *therefore it is considered as any region of the adaptor.*”) (emphasis added).

*sequence adjacent to flow-cell sequence 1 and the β SMI sequence adjacent to flow cell sequence 2.*²⁸⁶

The SMI sequence (α and β) is located on the adaptor in between sequences FC1 and FC2 sequences, respectively and the T-tailed DNA fragment. In other words, the SMI sequences (α and β) at a “junction between the adapter and the nucleic fragment.” This is consistent with the examiner’s interpretation of “junction,” as described *supra*. Thus, the ’389 PCT Appl. (Schmitt)’s description of each unique identifier being at a junction between the adapter and the nucleic fragment satisfies the claim.

11. Claim 12 of the ’241 Patent

179. Claim 12 recites “[t]he method of claim 1, further comprising determining the genomic start position for each sequencing read.” In my opinion, the ’389 PCT Appl. (Schmitt) discloses this claim. For example, the ’389 PCT Appl. (Schmitt)’s Example 1 aligns the sequencing reads to the reference human genome. “Reads were aligned to the human genome with the Burrows-Wheeler Aligner (BWA)”²⁸⁷

²⁸⁶ See Ex. 1007, ’389 PCT Appl. (Schmitt) at [0012] (emphases added).

²⁸⁷ *Id.* at [0096] (emphasis added); see also *id.* at step 10: “Align reads to the human genome.”

180. The Burrows-Wheeler Alignment tool (BWA) is a widely used algorithm to enable efficient mapping of short reads produced by NGS platform sequencers to a reference genome.

*We implemented Burrows-Wheeler Alignment tool (BWA), a new read alignment package that is based on backward search with Burrows–Wheeler Transform (BWT), to efficiently align short sequencing reads against a large reference sequence such as the human genome, allowing mismatches and gaps.*²⁸⁸

Li & Durbin (2009) explained that they “simulated reads from the human genome using the *wgsim program that is included in the SAMtools package* and ran the four programs [including BWA] to map the reads back to the human genome. *As we know the exact coordinate of each read*, we are able to calculate the alignment error rate.²⁸⁹ As described *supra*, SAMtools (just referenced) was widely used in the NGS community that also provided the genomic start position for each sequencing read. “(a) Alignments of one pair of reads and three single-end reads. (b) The corresponding SAM file. . . . *The coordinate shown in SAM is the position of the first aligned base.*”²⁹⁰ One of ordinary skill in the art would have known (i) that aligning a sequence read to a reference genome necessarily requires

²⁸⁸ See Ex. 1032, Li & Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform*, BIOINFORMATICS 25:1754-1760 (2009) (“Li & Durbin (2009)”) at Abstract.

²⁸⁹ *Id.* at 1758 (emphases added).

²⁹⁰ See Ex. 1011, Li (2009) at 2079 (emphasis added).

determining the genomic start positions of each read and (ii) that there were several alignment tools, such as BWA, or SAMtools²⁹¹ that provided reference genome coordinate positions of mapped sequence reads as part of their routine analysis. Thus, the '389 PCT Appl. (Schmitt)'s description of using widely used bioinformatic tools, such as BWA and SAMtools, to determine the genomic start position for each sequencing read satisfies the claim.

12. Claim 14 of the '241 Patent

181. Claim 14 recites “[t]he method of claim 1, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this claim. For example, the '389 PCT Appl. (Schmitt) discloses that a SMI adaptor may be a nucleotide n-mer sequence between 4 and 20 nucleotides in length.

Each n-mer sequence may be between approximately 3 to 20 nucleotides in length. *Therefore, each n-mer sequence may be approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides in length.*²⁹²

Example 1 of the '389 PCT Appl. (Schmitt) involves the use of “a 24-nucleotide SMI” that “*consists of two 12-nucleotide sequences* that can be designated XY.”²⁹³

²⁹¹ See Ex. 1007, '389 PCT Appl. (Schmitt) at [0096] for reference to BWA and [0103] for reference to SAMtools.

²⁹² *Id.* at [0026] (emphasis added).

²⁹³ *Id.* at [0100] (emphasis added).

Thus, the '389 PCT Appl. (Schmitt)'s description that the unique identifier can be at least 5, 6, 7, 8, 9, or 10 nucleotides in length satisfies the claim.

13. Claim 15 of the '241 Patent

182. Claim 15 recites “[t]he method of claim 1, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and performing an amplification reaction to obtain the amplicons.” In my opinion, the '389 PCT Appl. (Schmitt) discloses this claim. For example, the '389 PCT Appl. (Schmitt) discloses the following steps of a duplex consensus sequencing method to fragment DNA and ligated SMI adaptors to each end of the fragments.

Sheared double-stranded DNA that has been end-repaired and T-tailed is *combined with A-tailed SMI adaptors and ligated according to one embodiment*. . . . After size-selecting for appropriate length fragments, PCR amplification with primers containing Illumina flow-cell-compatible tails is carried out to generate families of PCR duplicates.”²⁹⁴

One of ordinary skill in the art would have understood that shearing a nucleic acid molecule means fragmenting that molecule. The '241 Patent also acknowledges that well known techniques of fragmenting DNA including shearing DNA.²⁹⁵ The

²⁹⁴ *Id.* (emphasis added).

²⁹⁵ See Ex. 1001 '241 Patent at 7:24-36. (“In some embodiments, *fragmentation of the nucleic acids can be achieved through methods known in the art.*”

'389 PCT Appl. (Schmitt) teaches other examples that meet this limitation as well.²⁹⁶

183. Example 1 of the '389 PCT Appl. (Schmitt) also discloses the claim steps of fragmenting the nucleic acid, appending the adaptors to the end of the fragments, and performing amplification to obtain the amplicons. Example 1 of the '389 PCT Appl. (Schmitt)'s method includes the following steps.

- **3 micrograms of DNA** was diluted into 130 microliters of TE buffer (10 mM tris-HCl, pH 8.0, 0.1 M EDTA) and **was sheared on the Covaris AFA system**²⁹⁷
- The custom adaptors were **ligated by combining 750 ng of T-tailed DNA with 250 pmol adaptors** in a reaction containing 3000 units T4 DNA ligase ²⁹⁸
- Pre-capture amplification. **375 ng adaptor-ligated DNA was PCR amplified** with primers AATGATACGGCGACCACCGAG (SEQ ID NO:3) and GTGACTGGAGTTCAGACGTGTGC (SEQ ID NO:4) using the Kappa high-fidelity PCR kit **for 8 cycles with an annealing temperature of 60 C.**²⁹⁹

Fragmentation can be achieved through physical fragmentation methods and/or enzymatic fragmentation methods. **Physical fragmentation methods can include** nebulization, sonication, and/or **hydrodynamic shearing. . . .** In some embodiments, **the fragmentation comprises treating the nucleic acids in the input sample with one or more enzymes** under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks.”) (emphases added).

²⁹⁶ See Ex. 1007, '389 PCT Appl. (Schmitt) at [0068] (using enzymes), (using shearing).

²⁹⁷ *Id.* at [0092] (emphases added).

²⁹⁸ *Id.* (emphasis added).

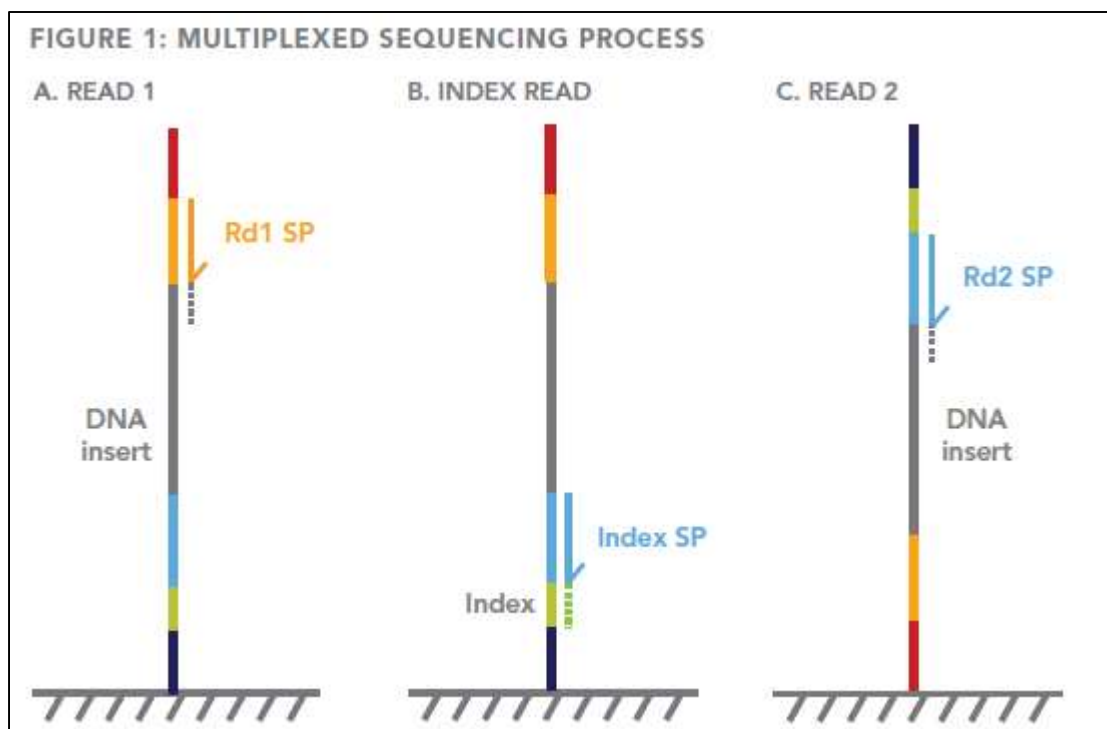
²⁹⁹ *Id.* at [0093] (emphases added).

Thus, the '389 PCT Appl. (Schmitt)'s disclosure of the steps of fragmenting the nucleic acid, appending the adaptors to the end of the fragments, and performing amplification to obtain the amplicons satisfies this claim.

D. Grounds IV and V: Claim 4 of the '241 Patent would have been obvious over the '202 Patent Pub. (Porreca) in view of Illumina (2008) and over '389 PCT Appl. (Schmitt) in view of Illumina (2008)

184. Claim 4 recites “[t]he method of claim 1, wherein the identifier site is sequenced separately from the target sequence.” In my opinion, one of ordinary skill in the art would have been motivated to sequence the identifier site separately from the target sequence in both the '202 Patent Pub. (Porreca)'s and '389 PCT Appl. (Schmitt)'s methods. One of ordinary skill in the art would have been aware of the advantages of performing separate sequencing steps. Typically, NGS platforms produce short reads, and due to this limitation, sequencing the identifier site separately from the target sequence results in a larger amount of target sequence, as the identifier site bases would not be included in that read. One of ordinary skill in the art would have understood that including more target sequence bases in a read would be beneficial, such as a longer sequence read can be more accurately mapped to a reference genome sequence than a short sequence read, and therefore, one of ordinary skill in the art would have thus been motivated to sequence the identifier sites separately, particularly with the Illumina platform that produces short sequence reads.

185. One of ordinary skill in the art also would have been motivated to apply the teachings of the Illumina platform in order to do so. For example, the Illumina platform since well before November 13, 2013, the priority date of the '241 Patent. As described in my 2010 review of NGS platform, I explained that “[c]urrently, the *Illumina/Solexa Genome Analyzer (GA) dominates the NGS market.*”³⁰⁰ One of ordinary skill in the art would have further known that the sequence adaptors in the Illumina platform were sequenced separately from the target sequence. For instance, Illumina (2008) shows that the target sequence is sequenced separately from the index sequence, as reproduced *infra*.³⁰¹



³⁰⁰ See Ex. 1010, Metzker (2010) at 34 (emphasis added).

³⁰¹ See Ex. 1008, Illumina (2008) at 1.

Illumina (2008) explains that “[s]ample multiplexing involves a total of three sequencing reads, including a separate index read.”³⁰² This demonstrates that one of ordinary skill in the art would have known how to sequence an adaptor sequence in a read separately from the target sequence as this was a routine function on the Illumina NGS platform. I understand that while Illumina (2008) demonstrates this with an index sequence, it would have been obvious for one of ordinary skill in the art to include an identifier sequence in the same location as the index sequence. In my opinion, one of ordinary skill in the art also would have been motivated to apply Illumina (2008)’s teachings to the ’202 Patent Pub. (Porreca) and ’389 PCT Appl. (Schmitt) because both references explicitly refer to the Illumina platform. The ’202 Patent Pub. (Porreca) teaches the following steps. “***Perform Sequencing of Library According to Manufacturer’s Directions (e.g. Illumina, ABI, etc).***”³⁰³ The ’389 PCT Appl. (Schmitt) also includes similar teachings. “***These products are then sequenced using any suitable method known in the art including,*** but not limited to, ***the Illumina sequencing platform, . . .***”³⁰⁴ The ’389 PCT Appl. (Schmitt) further explains that the “use of DCS ***does not require any significant deviations from the normal workflow of sample preparation for Illumina DNA***

³⁰² *Id.* at 1.

³⁰³ See Ex. 1006, ’202 Patent Pub. (Porreca) at [0295] (emphasis added).

³⁰⁴ See Ex. 1007, ’389 PCT Appl. (Schmitt) at [0074] (emphases added).

sequencing.”³⁰⁵ The ’389 PCT Appl. (Schmitt) additionally explains the advantages of having longer target sequencing reads, which would further motivate one of ordinary skill in the art to conduct the sequencing separately. “Combining information regarding the shear points of DNA with the SMI tag sequence would *allow a shorter SMI to be used, thus minimizing loss of sequencing capacity due to sequencing of the SMI itself.*”³⁰⁶

E. Grounds VI and VII: Claim 8 would have been obvious over the ’202 Patent Pub. (Porreca) in view of Illumina (2008) and over the ’389 PCT Appl. (Schmitt) in view of Illumina (2008)

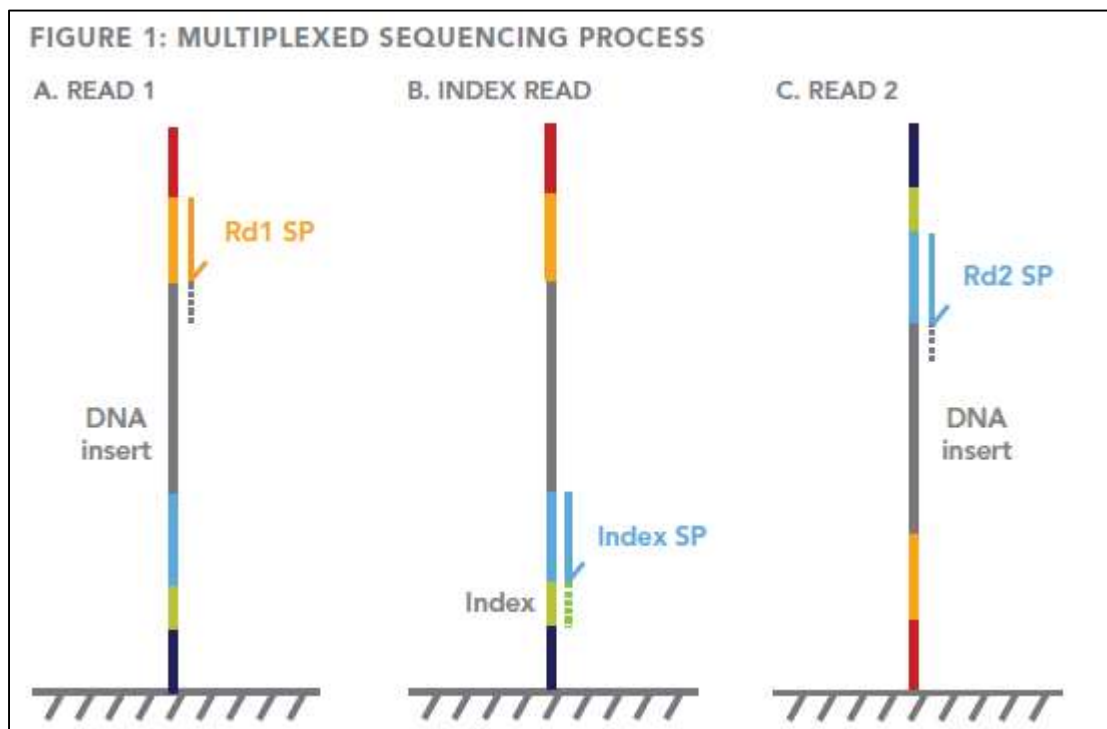
186. As I discuss *supra* in Grounds II and III, both the ’202 Patent Pub. (Porreca) and the ’389 PCT Appl. (Schmitt) disclose this claim. It also would have been obvious to one of ordinary skill in the art to include the claimed universal target sequencing primer binding site in both the ’202 Patent Pub. (Porreca) and the ’389 PCT Appl. (Schmitt) adaptors. One of ordinary skill in the art would have known that including such a binding site in the adaptor is necessary to carry out an NGS sequencing reaction.

187. For example, the Illumina platform produces millions, if not billions of clonally amplified template clusters on their flow cell after bridge amplification. To sequence each of template clusters, the only practical and most efficient manner

³⁰⁵ *Id.* at [0086] (emphasis added).

³⁰⁶ *Id.* at [0083] (emphasis added).

to do so is using universal target sequencing primer binding site, for which the Read 1 sequencing primer (“Rd1 SP”) and Read 2 sequencing primer (“Rd2 SP”) would hybridize to the binding site to start the SBS chemistry method of sequencing, as reproduced *infra*.³⁰⁷



One of ordinary skill in the art thus would have been motivated to use the claimed binding site into the '202 Patent Pub. (Porreca) and the '389 PCT Appl. (Schmitt adaptors) as both references are directed to NGS techniques, particularly the Illumina platform described *supra*.

³⁰⁷ See Ex. 1008, Illumina (2008) at 1.

188. Illumina (2008) instructs its user of the Illumina platform that “A: Application read 1 (dotted line) is generated *using the Read 1 Sequencing Primer (Rd1 SP)*. . . . the original strand is removed and the complementary strand acts as a template for application read 2 (dotted line), primed *by the Read 2 Sequencing Primer (Rd2 SP)*.”³⁰⁸ One of ordinary skill in the art thus would have known how to successfully include the claimed universal target sequencing primer binding site into the ’202 Patent Pub. (Porreca) and the ’389 PCT Appl. (Schmitt) adaptors.

189. One of ordinary skill in the art’s understanding that Illumina’s sequencing primer binding sites for Rd1 SP and Rd2 SP are universal target sequencing primer binding sites is further confirmed by Metzker (2010), an article that I authored that was published in the journal *Nature Reviews Genetics*. My review article discusses how the Illumina platform operates. “[A]daptors containing *universal priming sites* are ligated to the target ends, allowing complex genomes to be amplified with common PCR primers. . . . Solid-phase amplification can produce 100–200 million spatially separated template clusters (Illumina/Solexa), providing free ends to which a *universal sequencing primer* can be hybridized to initiate the NGS reaction.”³⁰⁹ The Illumina platform therefore provides a universal target sequencing primer binding site that was used in

³⁰⁸ *Id.* at Figure 1 legend (emphases added).

³⁰⁹ *See* Ex. 1010, Metzker (2010) at 32 (emphases added).

sequencing in the '202 Patent Pub. (Porreca) and the '389 PCT Appl. (Schmitt) methods.

190. For example, one of ordinary skill in the art also would have been motivated to apply Illumina (2008)'s teachings to the '202 Patent Pub. (Porreca) and the '389 PCT Appl. (Schmitt) because both references explicitly refer to the Illumina platform. For example, the '202 Patent Pub. (Porreca) teaches the following sequencing step. “***Perform Sequencing of Library According to Manufacturer’s Directions (e.g. Illumina, ABI, etc).***”³¹⁰ The '389 PCT Appl. (Schmitt) also provides a similar disclosure. “***These products are then sequenced using any suitable method known in the art including,*** but not limited to, ***the Illumina sequencing platform, ...***”³¹¹ The '389 PCT Appl. (Schmitt) further explains that the “use of DCS ***does not require any significant deviations from the normal workflow of sample preparation for Illumina DNA sequencing.***”³¹²

F. Ground VIII: Claim 12 would have been obvious over the '202 Patent Pub. (Porreca) in view of Li (2009)

191. Claim 12 recites “[t]he method of claim 1, further comprising determining the genomic start position for each sequencing read.” In my opinion, in view of the '202 Patent Pub. (Porreca) and Li (2009), it would have been obvious

³¹⁰ See Ex. 1006, '202 Patent Pub. (Porreca) at [0295] (emphasis added).

³¹¹ See Ex. 1007, '389 PCT Appl. (Schmitt) at [0074] (emphases added).

³¹² *Id.* at [0086] (emphasis added).

to one of ordinary skill in the art to have performed the method of claim 1 and to further determine the genomic start position for each sequencing read.

192. Li (2009) discloses the limitation “determining the genomic start position for each sequencing read.” Li (2009) describes an alignment format called, *Sequence Alignment/Map (SAM)* and discusses the “SAMtools” software package that ***provides universal tools for processing read alignments***.³¹³ SAMtools was a well-known tool that provides a common alignment format for NGS sequence reads mapped to a reference genome sequence.³¹⁴ The use of SAMtools was explicitly used in ’389 PCT Appl. (Schmitt). ***“PCR duplicates were filtered out with samtools rmdup, a standard tool which uses the shear points of DNA molecules to identify PCR duplicates,*** as molecules arising from duplicated DNA will have shared shear points.”³¹⁵

193. Li (2009)’s Figure 1 shows a Compact Ideosyncratic Gapped Alignment Report (CIGAR) and the corresponding SAM file in Figure 1(b). Li (2009) explains that ***“[t]he coordinate shown in SAM is the position of the first aligned base.*** The CIGAR string for this alignment contains a P (padding)

³¹³ See Ex. 1011, Li (2009) at 2078 (emphasis added).

³¹⁴ *Id.* (“SAMtools provides a ***“common alignment format that supports all sequence types and aligners creates a well-defined interface between alignment and downstream analyses,*** including variant detection, genotyping and assembly.”) (emphasis added).

³¹⁵ See Ex. 1007, ’389 PCT Appl. (Schmitt) at [0103] (emphasis added).

operation which correctly aligns the inserted sequences.”³¹⁶ One of ordinary skill in the art would have understood the Li (2009) disclosure to meet the limitation requiring “determining the genomic start position for each sequencing read.”

1. One of ordinary skill in the art would have been motivated to combine the '202 Patent Pub. (Porreca) with Li (2009) and would have had a reasonable expectation of success

194. In my opinion, one of ordinary skill in the art would have been motivated to combine the methods disclosed in the '202 Patent Pub. (Porreca) with the SAMtools alignment tool disclosed in Li (2009). SAMtools was a well-known tool that provides a common alignment format for NGS sequence reads mapped to a reference genome sequence.³¹⁷ As described *supra*, the '389 PCT Appl. (Schmitt) described SAMtools as being a “standard tool.” One of ordinary skill in the art would have been motivated to use the alignment and analysis techniques disclosed in Li (2009) because the '202 Patent Pub. (Porreca) is directed to NGS methods and SAMtools was commonly used in such analysis. For example, SAMtools was a

³¹⁶ See Ex. 1011, Li (2009) at 2079 (emphasis added).

³¹⁷ *Id.* (“SAMtools provides a “***common alignment format that supports all sequence types and aligners creates a well-defined interface between alignment and downstream analyses***, including variant detection, genotyping and assembly.”) (emphasis added).

well-known tool that provides a common alignment format for NGS sequence reads mapped to a reference genome sequence.³¹⁸

195. One of ordinary skill in the art also would have been motivated to combine the two references because SAMtools, and thus Li (2009), offers features that are beneficial to the '202 Patent Pub. (Porreca)'s methods. For example, the '202 Patent Pub. (Porreca) discloses that its invention is useful for determining biases from PCR.

Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are *useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification)*.³¹⁹

As described *supra* at Ground II, the '202 Patent Pub. (Porreca) discloses a method of detecting and removing PCR duplicates, which addresses amplification bias.

SAMtools also allows for the removal of PCR duplicates.

SAMtools is a library and software package for parsing and manipulating alignments in the SAM/BAM format. It is able to convert from other alignment formats, sort and merge alignments, *remove PCR duplicates, generate per-position*

³¹⁸ *Id.* (“SAMtools provides a “*common alignment format that supports all sequence types and aligners creates a well-defined interface between alignment and downstream analyses*, including variant detection, genotyping and assembly.”) (emphasis added).

³¹⁹ See Ex. 1006, '202 Patent Pub. (Porreca) at [0024] (emphases added).

information in the pileup format (Fig. 1c), . . . and show alignments in a text-based viewer.³²⁰

Both the '202 Patent Pub. (Porreca) and Li (2009) are thus clearly in the same field of endeavor and are complementary to one another in solving the same problem, *i.e.*, removing PCR duplicates from DNA sequencing reads to improve the accuracy of data analysis.

196. One of ordinary skill in the art also would have reasonably expected success in combining the '202 Patent Pub. (Porreca) and Li (2009). As described *supra*, SAMtools is a standard and well-known tool in the art, and one of ordinary skill in the art would have known how to use its features to carry out the methods disclosed in the '202 Patent Pub. (Porreca). For example, the '202 Patent Pub. (Porreca) created NGS libraries that were sequenced with an Illumina platform. “*After NGS reads have been generated, they are aligned to a known reference sequence or assembled de novo.*”³²¹ SAMtools was a well-known tool that provides a common alignment format for NGS sequence reads mapped to a reference genome sequence.³²² Thus, one of ordinary skill in the art also would have reasonably

³²⁰ See Ex. 1011, Li (2009) at 2079 (emphasis added).

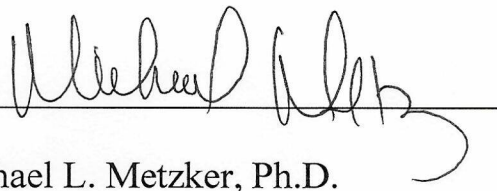
³²¹ See Ex. 1010, Metzker (2010) at 40 (emphasis added).

³²² See Ex. 1011, Li (2009) at 2078. (“SAMtools provides a “*common alignment format that supports all sequence types and aligners creates a well-defined interface between alignment and downstream analyses*, including variant detection, genotyping and assembly.”) (emphasis added).

expected success in combining the '202 Patent Pub. (Porreca) and Li (2009) to determine the genomic start position for each sequencing read.

I declare under penalty of perjury that the foregoing is true and correct. I declare that all statements made herein of my knowledge are true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: October 10, 2024

By: 
Michael L. Metzker, Ph.D.