UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

QIAGEN Sciences, LLC,
Petitioner,

v.

Tecan Genomics, Inc.,
Patent Owner.

Case IPR2025-00026

U.S. Patent No. 11,098,357

Declaration of Michael L. Metzker, Ph.D.

TABLE OF CONTENTS

Page

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 21
- E. The use of molecular identifiers to detect PCR duplicates has been well-known in the art 24

VI. APPLICABLE STANDARDS 25

- A. Relevant time period 25
- B. Level of ordinary skill in the art 26
- C. Anticipation 27
- D. Obviousness 28

VII. THE ’357 PATENT 30

VIII. CLAIM CONSTRUCTION 34

IX. PRIOR ART 35

- A. The ’202 Patent Pub. (Porreca) 35
- B. The ’389 PCT Appl. (Schmitt) 40
- C. The ’640 Patent Pub. (McCloskey) 43
- D. Illumina (2008) 48
- E. Metzker (2010) 50

X. ASSERTED GROUNDS 53

- A. Ground I: Claims 1-3, and 5-10 Are Anticipated by the ’202 Patent Pub. (Porreca) 53

B. Ground II: Claims 1-3, 5-10 Would Have Been Obvious Over the ’389 PCT Appl. (Schmitt) In View Of the ’640 Patent Pub. (McCloskey) 78

C. Grounds III and IV: Claim 4 of the ’357 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 ’640 Patent Pub. (McCloskey) in further view of Illumina (2008) 110

D. Grounds V and VI: Claim 7 would have been obvious over the ’202 Patent Pub. (Porreca) in view of Illumina (2008) and over ’389 PCT Appl. (Schmitt) in view of ’640 Patent Pub. (McCloskey) in further view of Illumina (2008) 114

QIAGEN Exhibit 1003.0003

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 '357 Patent.[1]

II. MATERIALS REVIEWED

2. I have reviewed the materials cited herein, including the following exhibits. The materials are publications that a person of ordinary skill[2] in the field 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, *Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR*, 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, *Multiplexed sequencing with the Illumina Genome Analyzer system* (2008) ("Illumina (2008)"), available at

[1] *See* Ex. 1002, Amorese *et al*., *Compositions and methods for identification of a duplicate sequencing read*, U.S. Patent No. 11,098,357 (the "'357 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 *Illumina 2008*
1010	Metzker, *Sequencing technologies – the next generation*, NATURE REVIEWS GENETICS 11:31-46 (2010) ("Metzker (2010)")
1011	Li *et al.*, *The Sequence Alignment/Map format and SAMtools*, 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 *et al.*, *Encoding PCR products with batch-stamps and barcodes*, BIOCHEMICAL GENETICS 45:761-767 (2007) ("McCloskey (2007)")
1016	Curriculum Vitae of Michael L. Metzker, Ph.D.
1017	Metzker & Caskey, *Polymerase chain reaction*, In ENCYCLOPEDIA OF MEDICAL DEVICES AND INSTRUMENTATION, Second Edition, Volume 5, pp. 380-387 (2006) ("Metzker (2006)")
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)")
1019	Saiki *et al.*, *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase*, SCIENCE 239:487-491 (1988) ("Saiki (1988)")
1020	Gibbs *et al.*, *Detection of single DNA base differences by competitive oligonucleotide priming*, NUCLEIC ACIDS RESEARCH 17:2437-2448 (1989) ("Gibbs (1989)")
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)")
1022	Chamberlain *et al.*, *Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification*, NUCLEIC ACIDS

Ex.	Description
	RESEARCH 16:11141-11156 (1988) ("Chamberlain (1989)")
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)")
1024	Grompe, *The rapid detection of unknown mutations in nucleic acids*, NATURE GENETICS 5:111-117 (1993) ("Grompe (1993)")
1025	Syvänen, *Accessing genetic variation: genotyping single nucleotide polymorphisms*, NATURE GENETICS 2:930-942 (2011) ("Syvänen (2001)")
1026	Metzker, *Emerging technologies in DNA sequencing*, GENOME RESEARCH 15:1767-1776 (2005) ("Metzker (2005)")
1027	Rothberg *et al*., *An integrated semiconductor device enabling non-optical genome sequencing*, NATURE 475:348-352 (2011) ("Rothberg (2011)")
1028	Jain *et al*., *Improved data analysis for the MinION nanopore sequencer*, NATURE METHODS 12:351-356 (2015) ("Jain (2015)")
1029	Illumina, *Multiplexing, Sample preparation guide* (2010) ("Illumina Manual (2010)")
1030	Illumina, *TruSeq™ RNA and DNA sample preparation kits* (2010) ("Illumina TruSeq Data Sheet (2010)")
1031	Craig *et al*., *Identification of genetic variants using barcoded multiplexed sequencing*, NATURE METHODS 5:887-893 (2008) ("Craig (2008)")
1032	Li & Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform*, BIOINFORMATICS 25:1754-1760 (2009) ("Li & Durbin (2009)")
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)")

III. SUMMARY OF OPINIONS

3. In my opinion, Claims 1-10 of the '357 Patent are unpatentable as anticipated and/or obvious. My opinion is based on the following grounds:

a. Ground 1: Claims 1-3, 5-10 are anticipated under 35 U.S.C. § 102(a) by the '202 Patent Pub. (Porreca).[3]

b. Ground 2: Claims 1-3, 5-10 would have been obvious under 35 U.S.C. § 103 over the '389 PCT Appl. (Schmitt)[4] in view of the '640 Patent Pub. (McCloskey).[5]

c. Ground 3: Claim 4 would have been obvious under 35 U.S.C. § 103 over the '202 Patent Pub. (Porreca) in view of Illumina (2008)[6]

d. Ground 4: Claim 4 would have been obvious under 35 U.S.C. § 103 over the '389 PCT Appl. (Schmitt) in view of the '640 Patent Pub. (McCloskey) in further view of Illumina (2008)

e. Ground 5: Claim 7 would have been obvious under 35 U.S.C. § 103 over the '202 Patent Pub. (Porreca) in view of Illumina (2008)

---

[3] *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)").

[4] *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)").

[5] *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)").

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

f. Ground 6: Claim 7 would have been obvious under 35 U.S.C. § 103 over the ’389 PCT Appl. (Schmitt) in view of the ’640 Patent Pub. (McCloskey) in further view of Illumina (2008)

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.

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.[7]

[7] *See* Ex. 1016, Metzker CV.

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 '357 Patent is related to the field of "high throughput sequencing reactions" and discloses methods of detecting duplicate nucleic acid sequencing reads.[8] 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 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

[8] *See* Ex. 1002, '357 Patent at 1:26-29.

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.[9] 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'-

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

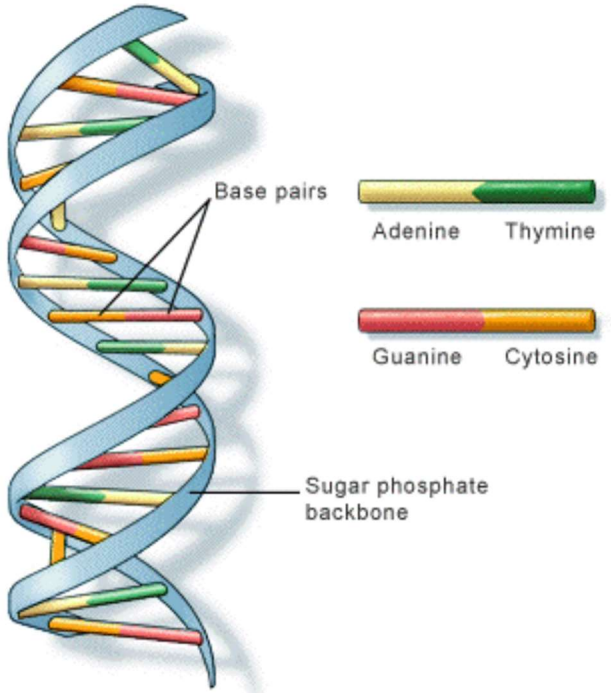
deoxyribonucleoside-5'-triphosphates or "dNTPs," and those added during RNA synthesis are called ribonucleotide-5'-triphosphates or "NTPs" or sometimes "rNTPs."

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 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*.



U.S. National Library of Medicine

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

QIAGEN Exhibit 1003.0019

create one or more copies (also called *amplification*), such as the polymerase chain reaction ("PCR").[10]

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 was invented by Dr. Kary Mullis in the 1980s.[11] PCR is used to specifically amplify a small amount of nucleic acid molecules, generating thousands to millions of copies of the target nucleic acid.[12]

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

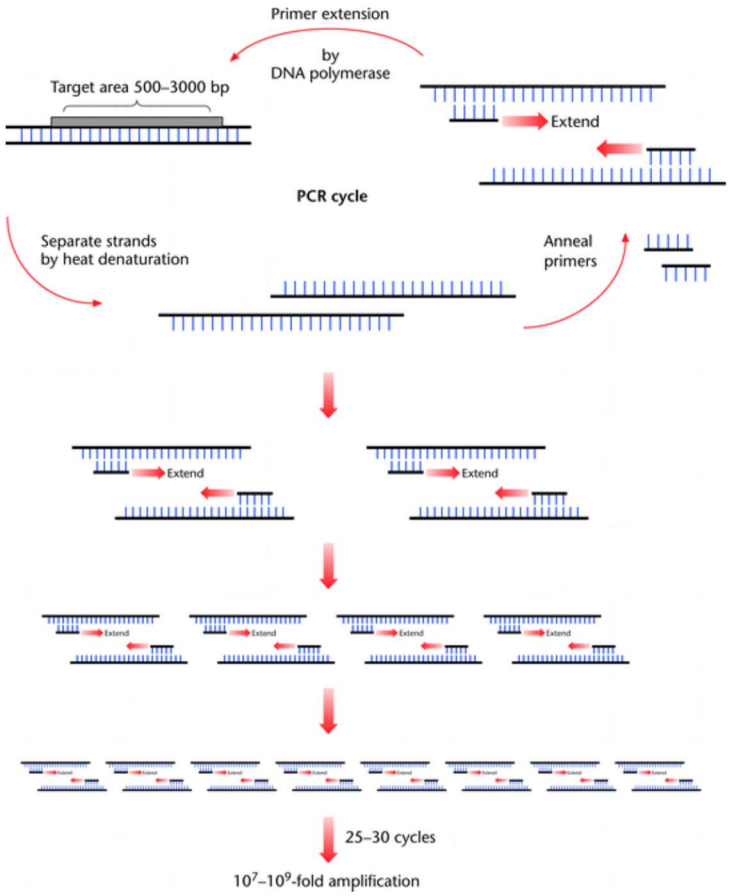
---

[10] *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)").

[11] *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.

[12] *See* generally Ex. 1018, Saiki (1985) and Ex. 1019, Saiki (1988).

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 template strand in PCR that define the targeted sequence of interest. The general concept of PCR is illustrated by the diagram *infra*.[13]



[13] *See* Ex. 1017, Metzker (2006) at Figure 1.

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 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.[14]

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.

---

[14] *Id*. at 380. ("Each cycle doubles the region marked by the primer sequences.").

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 priming[15] or called *amplification refractory mutation system*[16] 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 dystrophy[17] and Lesch—Nyhan syndrome.[18] 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

---

[15] *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)").

[16] *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)").

[17] *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)").

[18] *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)").

methods that involve scanning methods (*i.e.*, techniques that do not involve DNA sequencing) of PCR products for the presence of genetic alterations[19], genotyping methods, such as *mini-sequencing*, of PCR products to determine single nucleotide variations ("SNVs"),[20] 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.[21] 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 *infra* 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

---

[19] *See* Ex. 1024, Grompe, *The rapid detection of unknown mutations in nucleic acids*, NATURE GENETICS 5:111-117 (1993) ("Grompe (1993)").

[20] *See* Ex. 1025, Syvänen, *Accessing genetic variation: genotyping single nucleotide polymorphisms*, NATURE REVIEW GENETICS 2:930-942 (2011) ("Syvänen (2001)").

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

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.

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,[22] 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.[23] 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,[24] Life Technologies—the SOLiD® instrument,[25] and Pacific Biosciences—the single-molecule real-time or SMRT®

---

[22] *See* Ex. 1026, Metzker, *Emerging technologies in DNA sequencing*, GENOME RESEARCH 15:1767-1776 (2005) ("Metzker (2005)") at 1768-1770.

[23] *See, e.g.,* Ex. 1010, Metzker, *Sequencing technologies — the next generation*, NATURE REVIEWS GENETICS 11:31-46 (2010) ("Metzker (2010)").

[24] *Id.* at 39, Figures 1a, 3c, 3d.

[25] *Id.* at 36-39 at Figures 1a, 3a, 3b.

instrument.[26] Other NGS systems that have been commercialized include the Ion Torrent—the Personal Genome Machine ("PGM®")[27] and Ion Proton® instruments and Oxford Nanopore—the MinION sequencer.[28] 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.[29] To create genomic libraries, the DNA is typically sheared or enzymatically digested into smaller fragments, and platform-specific oligonucleotides, called adaptors[30] are

---

[26] *Id.* at 39 at Figures 1e, 4.

[27] *See* Ex. 1027, Rothberg *et al.*, *An integrated semiconductor device enabling non-optical genome sequencing*, NATURE 475:348-352 (2011) ("Rothberg (2011)").

[28] *See* Ex. 1028, Jain *et al.*, *Improved data analysis for the MinION nanopore sequencer*, NATURE METHODS 12:351-356 (2015) ("Jain (2015)").

[29] *See* Ex. 1010, Metzker (2010) at 32.

[30] *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.

attached or ligated to both ends of the DNA fragment.[31] These fragments can be 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 '357 Patent.[32] By that time, in fact, standard kits to prepare sequencing libraries had been available for at least several years before the '357 Patent. These kits, like the asserted claims of the '357 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 '357 Patent,[33] the use of molecular identifiers to identify PCR duplicates was well

---

[31] *See* Ex. 1010, Metzker (2010) at 32.

[32] *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)").

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

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*** prior to PCR amplification."[34] 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."[35] Miner (2004)'s method allowed for the "confirmation of the independent genomic origin of all sequences in final data sets derived from PCR amplification,"[36] 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 '357 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*.

A. Relevant time period

---

[34] *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).

[35] *Id*. at 1.

[36] *Id*.

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 '357 Patent.[37]

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 library preparation and sequencing experience, including but not limited to high-

[37] For this IPR proceeding, I am not offering an opinion whether the '357 Patent is entitled to the priority date of November 13, 2013.

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 skilled 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 '357 PATENT

47. The '357 Patent is titled, *Compositions and Methods for Identification of a Duplicate Sequencing Read*. It issued from Application No. 16/743,604 (the "'604 application"),[38] which was filed on January 15, 2020. The '357 Patent claims priority to U.S. Provisional Application No. 61/903,826 (the "'826 Provisional"), which was filed on November 13, 2013.

48. The '357 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."[39] The '357 Patent states that "[i]n RNA sequencing applications, accurate gene expression measurements may be hampered by PCR duplicate artifacts that occur during library

[38] *See* Ex. 1002, '357 Patent at cover page.

[39] *Id.* at 1:27-30.

amplification."[40] 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."[41] The '357 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."[42]

49. The '357 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."[43] To identify duplicates, the patent discloses the use of an "identifier site" that can be included in an "adaptor" appended to the target nucleotide to be sequenced.[44] The '357 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 polynucleotide strand of

---

[40] *Id.* at 1:35-37.

[41] *Id.* at 1:37-42.

[42] *Id.* at 1:42-46.

[43] *Id.* at 2:7-9.

[44] *Id.* at 4:19-38.

interest."[45] An identifier site "is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads."[46] An example of the application of an adaptor containing an identifier site is shown in the '357 Patent's Figure 2A, as illustrated *infra*:[47]

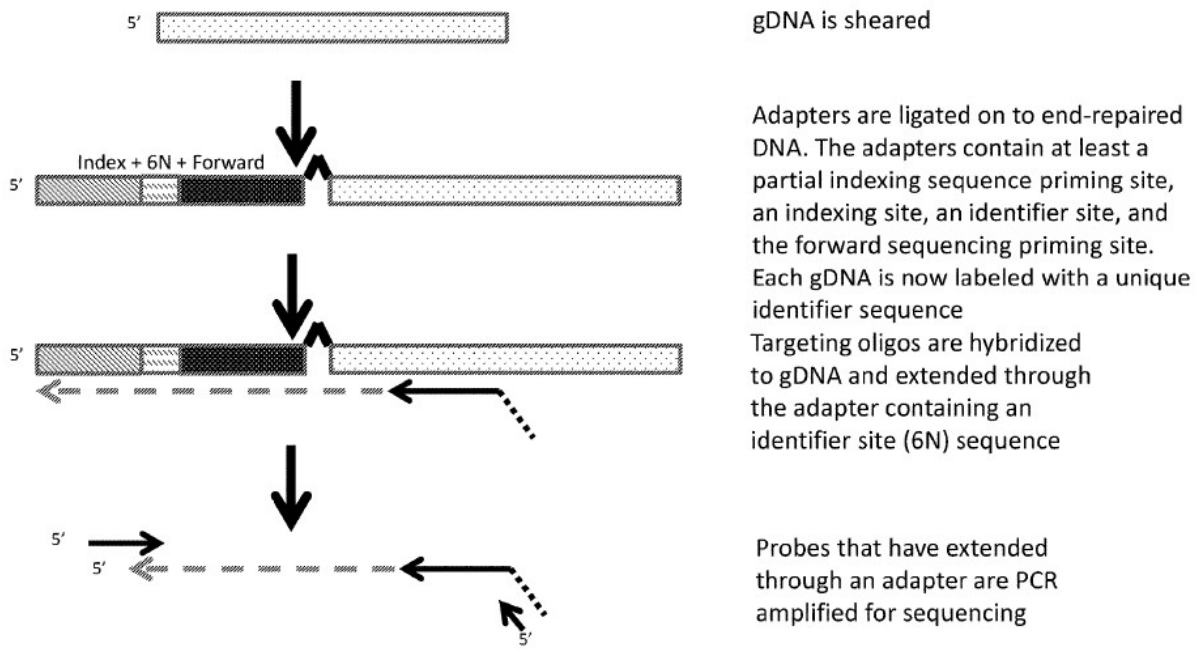


Figure 2A

50. As shown *supra*, Figure 2A shows an adaptor containing "at least a partial indexing sequence priming site, an indexing site, an identifier site, and the

---

[45] *Id.* at 4:20-25.

[46] *Id.* at 4:49-51.

[47] *Id.* at Figure 2A.

forward sequencing priming site" that is "ligated on to end-repaired DNA."[48] The '357 Patent further states that "a sequencing read is identified as a duplicate sequencing read when it contains an identifier site and target sequence that is the same as another sequencing read from the same population of sequencing reads."[49] The '357 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 '357 Patent recite methods for identifying PCR duplicates. Claim 1 is reproduced *infra*.

> The invention claimed is:
>
> 1. A method for detecting duplicate sequencing reads, the method comprising:
>
> ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides,
>
> an indexing site unique to a subset of the adaptors, and
>
> a primer binding site;

---

[48] *Id.*

[49] *Id.* at 14:16-20.

[50] *Id.* at 14:63-67.

amplifying the adaptor-ligated fragments into amplicons;

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

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

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 '357 Patent and understand that the patent 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."[51] I am also aware that during prosecution, 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."[52]

53. I further understand that the specification states that "appending the adaptor to the nucleic acid fragments generated by methods described herein can be

---

[51] *Id.* at 4:20-25.

[52] *See* Ex. 1013, '357 Patent File History at 0077.

achieved using a ligation reaction or a priming reaction."[53] An adaptor according to the '357 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 '357 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 the plain and ordinary meaning.

IX. PRIOR ART

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

A. The '202 Patent Pub. (Porreca)

55. The '202 Patent Pub. (Porreca) is a U.S. Patent Application Publication titled, *Methods and compositions for evaluating genetic markers*.[54] The '202 Patent Pub. (Porreca) was filed on March 13, 2012 and published on June 28, 2012.[55] 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 regions of interest. The '202 Patent Pub. (Porreca) discloses "methods and compositions that

---

[53] *See* Ex. 1002, '357 Patent at 9:43-46.

[54] *See* Ex. 1006, '202 Patent Pub. (Porreca) at cover page.

[55] *Id*.

are useful to reduce bias and increase the reproducibility of multiplex analysis of genetic loci."[56]

56. 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."[57] 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."[58] 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 to bias in a***

---

[56] *Id*. at Abstract.

[57] *Id*. at [0008] (emphasis added).

[58] *Id*. at [0023].

preparative step, for example, amplification)."[59] 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.

57. The '202 Patent Pub. (Porreca) teaches several methods of incorporating the differentiator tag.[60] For example, "nucleic acids comprising differentiator tag sequences may be incorporated by ligation."[61] "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."[62] 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)."[63] 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*.

---

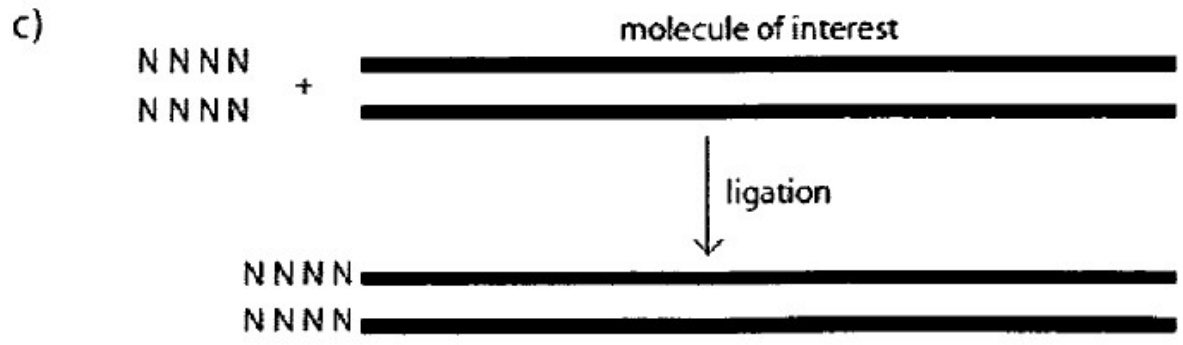
[59] *Id*. at [0024] (emphases added).

[60] *Id*. at [0150].

[61] *Id*.

[62] *Id*. (emphasis added).

[63] *Id*.



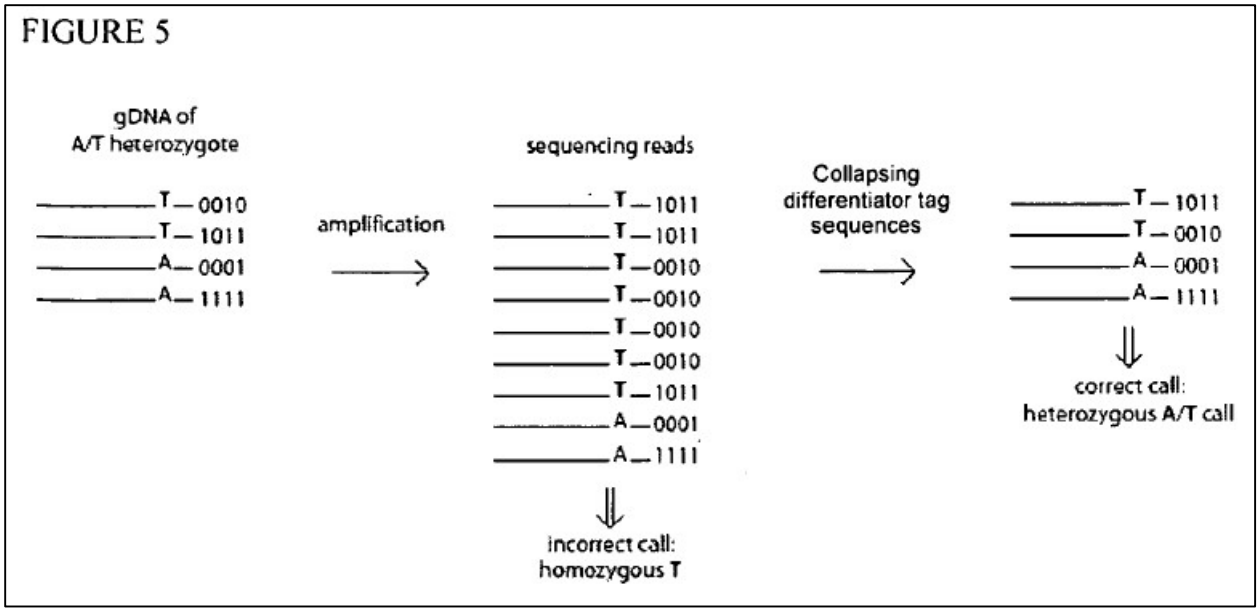
The ’202 Patent Pub. (Porreca) also explains that “***FIG. 4***A-***C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences***.”[64]

58. 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***.”[65] Figure 5 of the ’202 Patent Pub. (Porreca) depicts this process, as illustrated *infra*.

---

[64] *Id*. at [0140] (emphases added).

[65] *Id*. at [0298] (emphasis added).



59. 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 '357 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."[66] 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."[67]

B. The '389 PCT Appl. (Schmitt)

60. 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*.[68] It was filed on March 15, 2013 and was published on September 26, 2013.[69] 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").

61. 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

---

[66] *See* Ex. 1002, '357 Patent at 1:42-45 (emphasis added).

[67] *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).

[68] *See* Ex. 1007,'389 PCT Appl. (Schmitt) at cover page.

[69] *Id*.

fragments being sequenced and that uniquely identifies each fragment.[70] This SMI adaptor molecule is shown in Figure 1 of the '389 PCT Appl. (Schmitt), as illustrated *infra*.

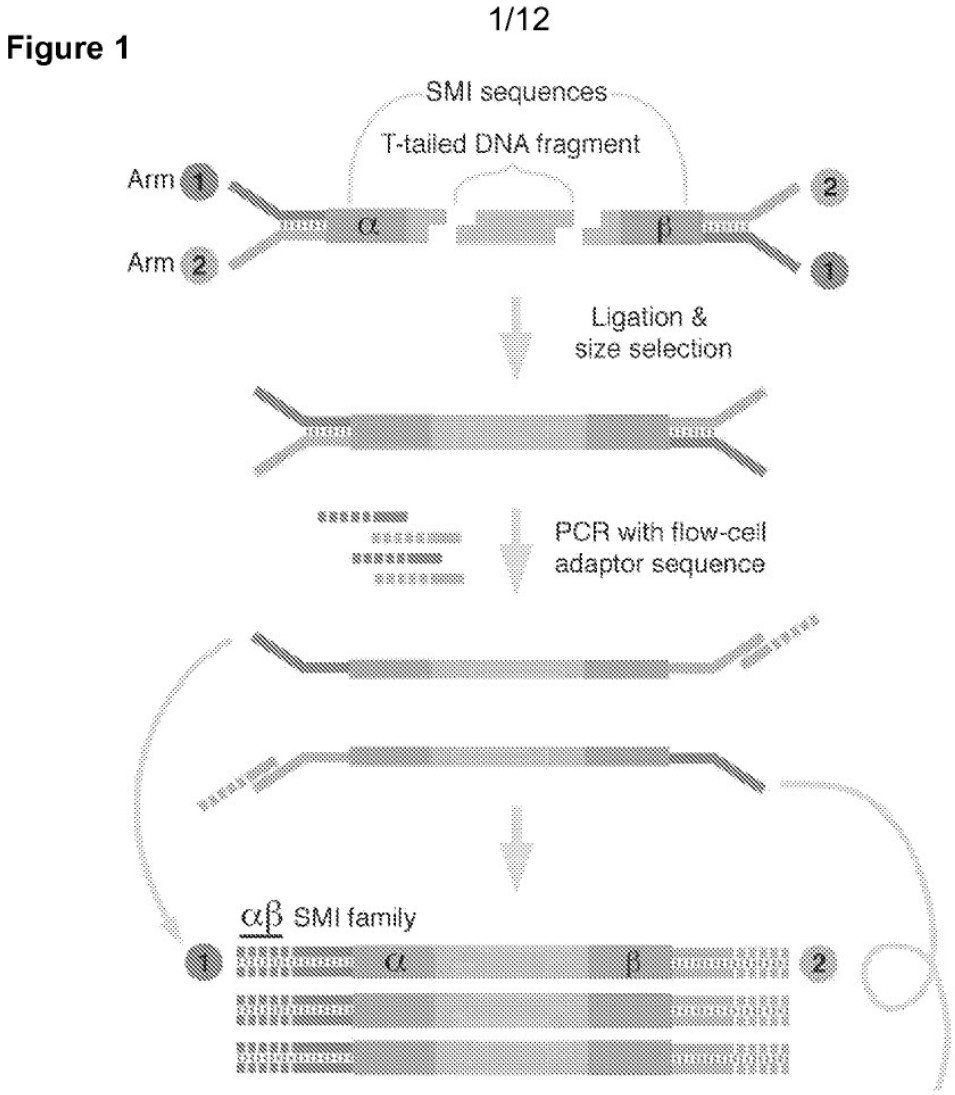


Figure 1 of the '389 PCT Appl. (Schmitt) shows that an adaptor containing an SMI sequence (represented as α and β segments) is ligated to the target sequence. The

[70] *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.”[71] The resulting amplified products are then sequenced, as illustrated *supra* in the bottom portion of Figure 1 of the ’389 PCT Appl. (Schmitt).

62. 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***.”[72] Because it is desirable to “restrict sequencing analysis to uniquely sequenced DNA fragments, ***many sequencing studies include a step to filter out PCR duplicates***.”[73] 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.”[74] As described *supra*, one of ordinary skill in the art would have understood that collapsing identical sequences having common

---

[71] *Id*. at [0012].

[72] *Id*. at [0085] (emphasis added).

[73] *Id*. (emphasis added).

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

target and SMI sequences in a single consensus read would require discarding or removing the other sequences in that group. I note that the '357 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***."[75]

C. The '640 Patent Pub. (McCloskey)

63. 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*.[76] It was filed on July 21, 2005 and published on January 25, 2007.[77] 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,"[78] which allows for the identification of duplicates as well as the sample from which the DNA originated.

64. 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***

---

[75] *See* Ex. 1002, '357 Patent at 1:42-45 (emphasis added).

[76] *See* Ex. 1005,'640 Patent Pub. (McCloskey) at cover page.

[77] *Id*.

[78] *Id*. at Abstract.

degree of redundancy for a small portion of the original genomic sequences."[79] 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.[80] 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.[81] The barcoding allows for the identification of redundant sequences,[82] and the batch-stamp identifies which sample each sequence originated

---

[79] *Id*. at [0003] (emphasis added).

[80] *Id*. at [0106]. "Here, a similar concept was applied to ***the labeling of individual genomic fragments with distinct sequence tags. The ability to bar-code 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).

[81] *Id*. at [0060].

[82] *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 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).

from.[83] The ’640 Patent Pub. (McCloskey) further teaches the removal of duplicate reads.[84]

65. 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*.

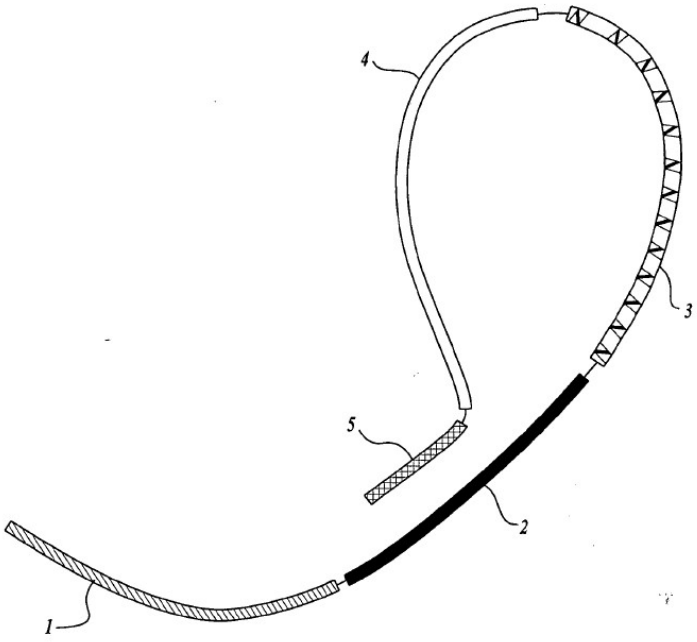


Fig.1.

[83] *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).

[84] *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).

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.”[85] Sequence (1) hybridizes to the target nucleic acid strand. DNA polymerase then extends sequence (1) along the target strand,[86] generating a complete strand that is complementary to the target sequence and that is attached to 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.[87] 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***.”[88] 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

---

[85] *Id*. at [0013].

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

[87] *Id*., *e.g.*, at [0056]-[0061].

[88] *Id*., *e.g.*, at [0063].

batch-stamp.[89] One of ordinary skill in the art would have understood that a hairpin linker is commonly used as an adapter sequence.

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

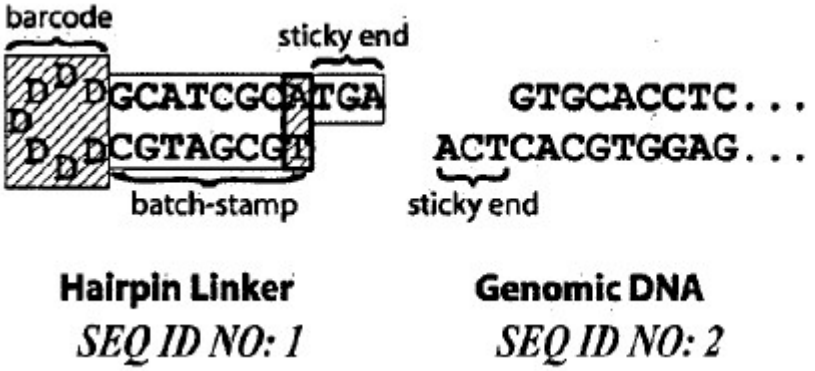


Fig.3.

Example 6 in the ’640 Patent Pub. (McCloskey) describes the use of the hairpin adaptor to identify redundant or duplicate sequences.[90] 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.[91] 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***

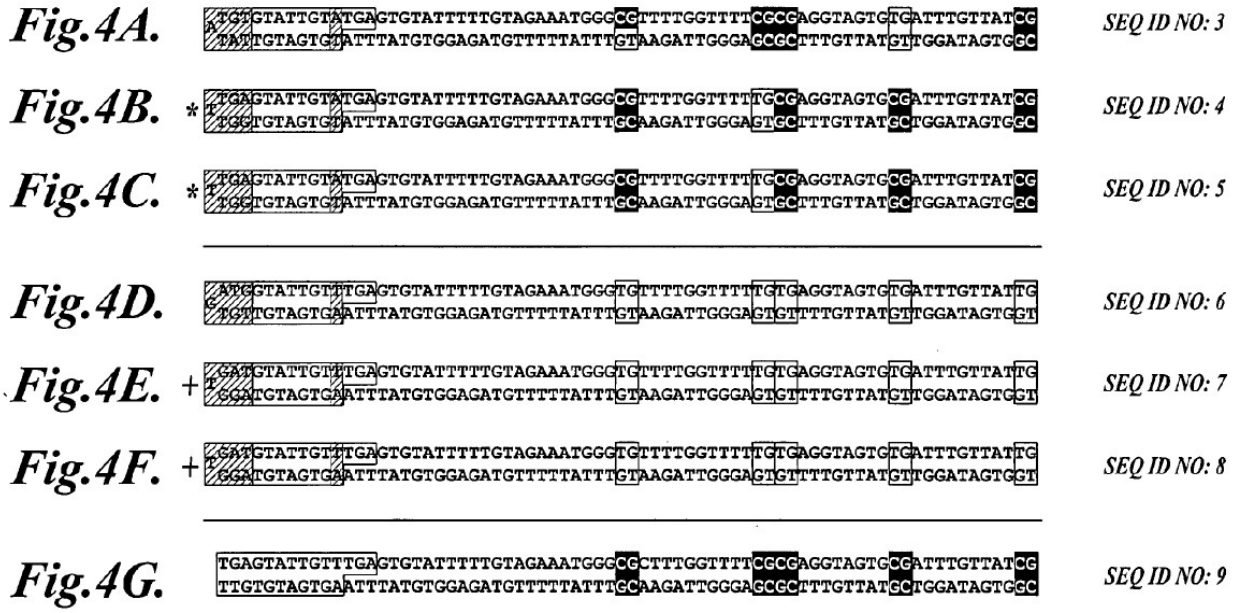
---

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

[90] *Id*.

[91] *Id*. at [0095].

these molecular labels in the hairpin linker sequence (FIG. 3) that is used in hairpin-bisulfite PCR."[92] The resulting product is amplified by PCR and then sequenced.[93] The sequencing results include reads of the barcode, batch-stamp, and target sequences, and are shown *infra* in Figures 4A-G of the '640 Patent Pub. (McCloskey).[94]



D. Illumina (2008)

67. Illumina (2008) is a datasheet published by Illumina titled, *Multiplexed sequencing with the Illumina Genome Analyzer system*.[95] The Illumina (2008)

---

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

[93] *Id*. at [0098], [0103].

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

[95] *See* Ex. 1008, Illumina (2008) at p. 1.

datasheet is dated December 2, 2008.[96] I have been informed by counsel and understand that Illumina (2008) is prior art to the '241 Patent under 35 U.S.C. 102(a). This reference explains Illumina's multiplex sequencing process, as illustrated *infra*.

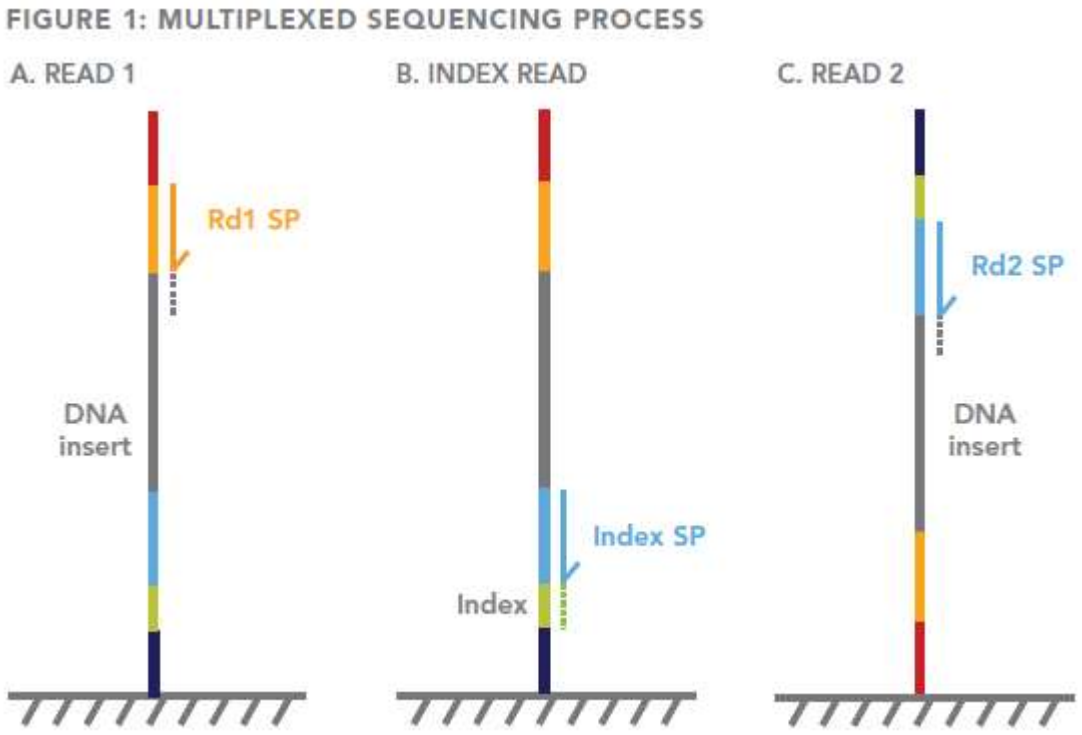


Figure 1 of Illumina (2008) explains "[s]ample multiplexing involves a total of three sequencing reads, ***including a separate index read, which is generated automatically on the Genome Analyzer equipped with the Paired-End Module***."[97] For the left-hand side of the figure *supra*, that is "A: Read 1," Illumina (2008) explains that "A: Application read 1 (dotted line) is generated using the Read 1

[96] *Id*. at p. 4.

[97] *Id*. at Figure 1 legend.

Sequencing Primer (Rd1 SP)."[98] 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)***."[99] 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."[100]

E. Metzker (2010)

68. Metzker (2010) is an article I authored that was published in the journal *Nature Reviews Genetics*, and that is titled, *Sequencing technologies - the next generation*.[101] Metzker (2010) was published online on December 8, 2009[102] and is "a technical review of template preparation, sequencing and imaging, genome alignment and assembly approaches, and recent advances in current and near-term commercially available NGS instruments."[103] My review article focused on technologies from "Roche/454, Illumina/Solexa, Life/APG and Helicos

---

[98] *Id*.

[99] *Id*. (emphasis added).

[100] *Id*.

[101] *See* Ex. 1010, Metzker (2010) at 31.

[102] *Id*.

[103] *Id*. at Abstract.

BioSciences, the Polonator instrument and the near-term technology of Pacific Biosciences."[104] 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."[105]

69. 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."[106] 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 template clusters (Illumina/Solexa), providing free ends to which a universal sequencing

---

[104] *Id*. at 31.

[105] *Id*.

[106] *Id*. at 32.

primer can be hybridized to initiate the NGS reaction."[107] Figure 1b of Metzker (2010) demonstrates this process, as illustrated *infra*.

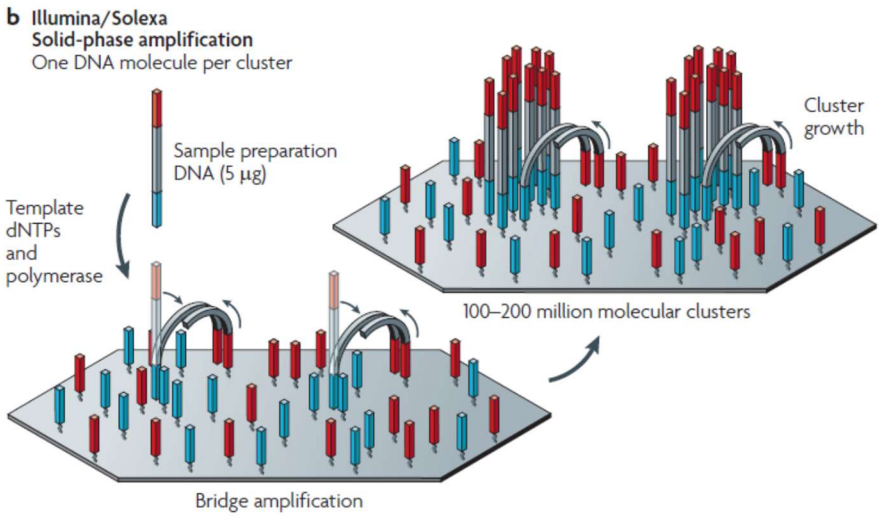


Figure 1b of Metzker (2010) explains that "[s]olid-phase amplification (b) is composed of two basic steps: ***initial priming and extending*** of the single-stranded, single-molecule template, and ***bridge amplification*** of the immobilized template with immediately adjacent primers to form clusters."[108] These bridge amplified templates formed "molecular clusters" which then sequenced using the Illumina SBS chemistry to generate sequence reads,[109] that could then be analyzed using SAMtools to identify duplicate sequence reads.

---

[107] *Id*.

[108] *Id*. at Figure 1 legend (emphasis added).

[109] *Id*. at 34-36, Figure 2a.

X. ASSERTED GROUNDS

A. Ground I: Claims 1-3, and 5-10 Are Anticipated by the ’202 Patent Pub. (Porreca)

70. In my opinion, *Porreca* anticipates claims 1-3, and 5-10 of the ’357 Patent. Below is a table matching the claim terms in the ’357 Patent to the corresponding disclosure in the ’202 Patent Pub. (Porreca).

’241 patent	**’202 Patent Pub. (Porreca)**
“identifier site”	“differentiator tag sequence”
“indexing site”	“barcode”

1. Independent Claim 1 of the ’357 Patent

(a) Limitation 1.p

71. 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."[110] The '202 Patent Pub. (Porreca) also explains that "***each nucleic acid molecule may be tagged with a unique differentiator tag***."[111] 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***)."[112]

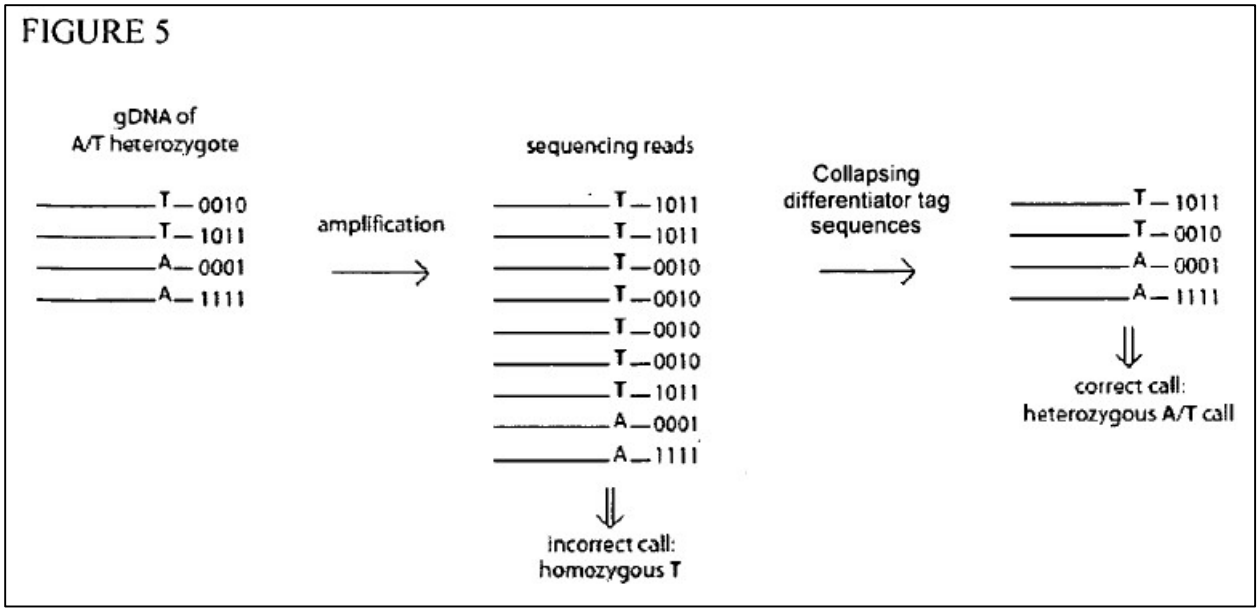
72. The '202 Patent Pub. (Porreca)'s differentiator tags are attached to the nucleic acid fragments prior to amplification, as illustrated *infra* in Figure 5.

---

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

[111] *Id*. at [0142] (emphasis added).

[112] *Id*. at [0024] (emphases added).



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.

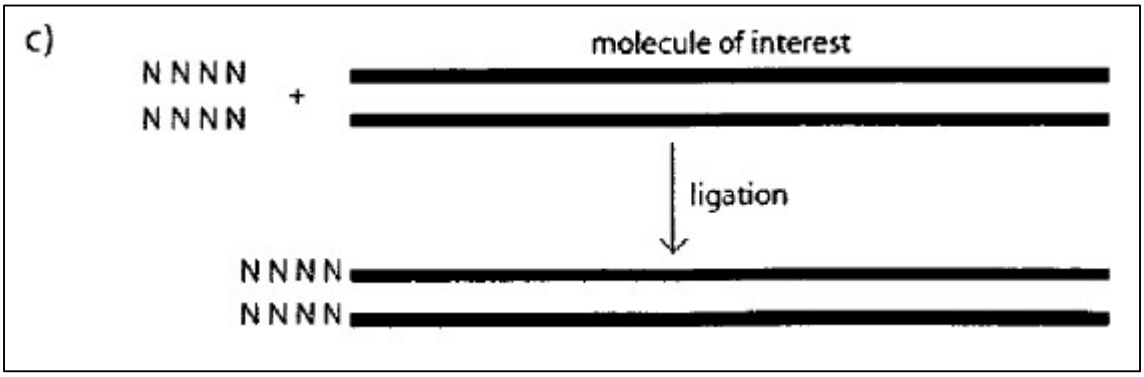
(b) Limitation 1.a

73. Limitation 1.a recites "ligating an adaptor to each of a plurality of nucleic acid fragments." 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 through a ligation 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***."[113]

74. Figure 4C illustrates an example of this, where an adaptor comprising a differentiator tag sequence (NNNN) is ligated to the nucleic acid molecule of interest, as reproduced *infra*.[114]



The '202 Patent Pub. (Porreca) also explains that "***FIG. 4***A-***C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences***."[115] This adaptor can include other sequences as well, such as a barcode to identify the sample of origin in a multiplex reaction, and a primer binding site, as explained *infra* in Limitations 1.c and 1.d.

---

[113] *Id.* at [0150].

[114] *Id.* at Fig. 4c.

[115] *Id*. at [0140] (emphases added).

(c) Limitation 1.b

75. Limitation 1.b recites "wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides." 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 '357 Patent explains that an "identifier site" is "a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads."[116] 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."[117] 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."[118] 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)

---

[116] *See* Ex. 1002, '357 Patent at 4:49-51.

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

[118] *Id*. at [0141] (emphasis added).

discloses a sufficient number of identifier sequences to uniquely tag up to 1.125×10^{15} nucleic acid fragments.

76. The ’202 Patent Pub. (Porreca) also discloses that each nucleic acid molecule is uniquely tagged with its own differentiator tag sequence. For example, ’202 Patent Pub. (Porreca) explains “***the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules***.”[119] 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*** . . . .”[120] 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.”[121]

(d) Limitation 1.c

77. Limitation 1.c recites “an indexing site unique to a subset of the adaptors.” In my opinion, the ’202 Patent Pub. (Porreca) discloses this claim. For example, the ’202 Patent Pub. (Porreca) teaches adaptors may include “barcodes” to uniquely identify particular samples.

---

[119] *Id*. at [0023] (emphasis added).

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

[121] *Id*. at [0139].

> 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 /0020640Al (McCloskey et al.).[122]

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."

78. 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***"[123] The '202 Patent Pub. (Porreca) also explains that "[i]n 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***."[124]

---

[122] *Id*. at [0149] (emphasis added).

[123] *Id*. at Abstract (emphasis added).

[124] *Id*. at [0102] (emphasis added).

79. 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.

(e) Limitation 1.d

80. Limitation 1.d recites "a primer binding site." In my opinion, the '202 Patent Pub. (Porreca) discloses this limitation. 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***."[125] 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***."[126]

---

[125] *Id*. at [0149] (emphasis added).

[126] *Id*. at [0150] (emphases added).

81. 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 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.[127]

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

(f) Limitation 1.e

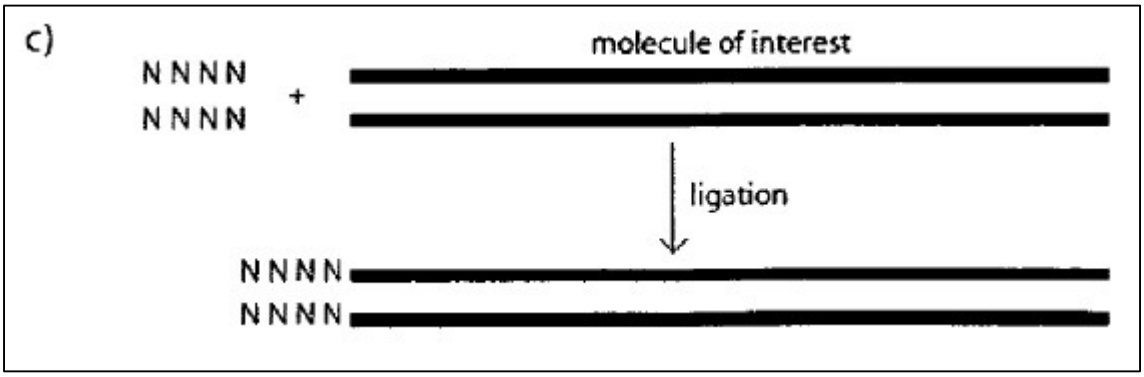
82. Limitation 1.e recites “amplifying the adaptor-ligated fragments into amplicons.” In my opinion, the ’202 Patent Pub. (Porreca) discloses this limitation. For example, the ’202 Patent Pub. (Porreca) teaches that an adaptor containing a differentiator tag sequence may be appended to a nucleic acid sequence of interest, either through a ligation reaction or by a primer extension reaction.

---

[127] *Id*. (emphases added).

> 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.[128]

83. 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*.



[128] *Id*. at [0150] (emphases added).

The ’202 Patent Pub. (Porreca) also explains that “***FIG. 4*A-*C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences***.”[129] Once the adaptor comprising a differentiator tag is appended 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 an amplicons, as reproduced *infra*.

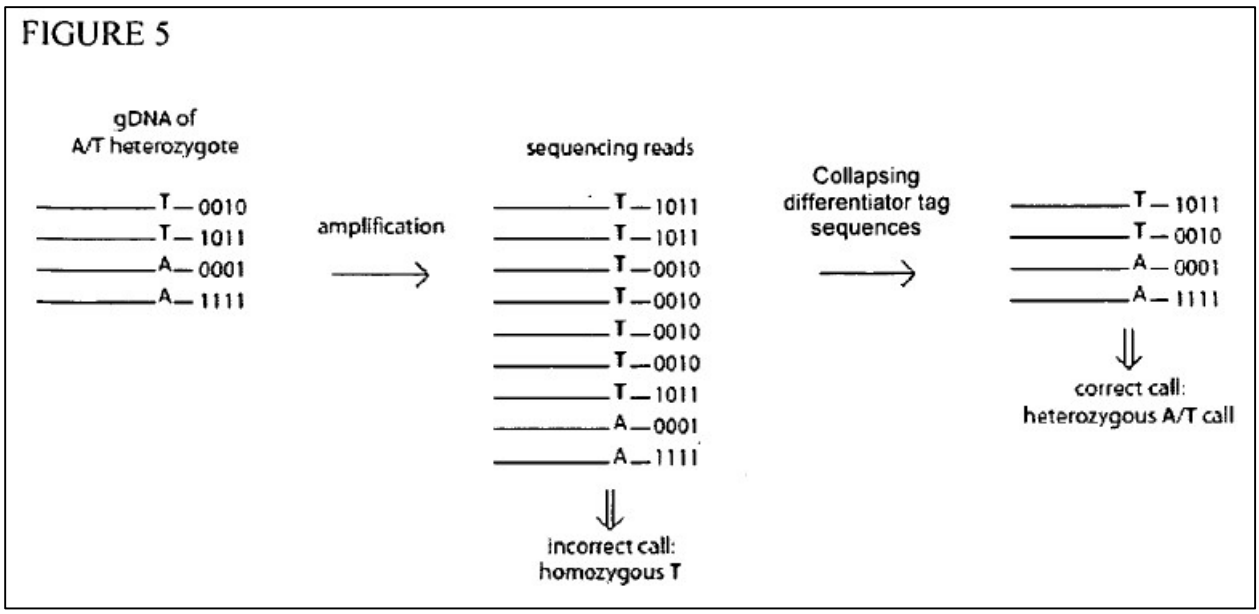


Figure 5 of the ’202 Patent Pub. (Porreca) shows four nucleic acid molecules, each having an appended adaptor including a differentiator tag sequence (*e.g.*, “0010” or “1011”). These adaptor-appended nucleic acid molecules undergo “amplification”

[129] *Id*. at [0140] (emphases added).

to produce several amplicons. This is shown under "sequencing reads" in the middle column of Figure 5 of the '202 Patent Pub. (Porreca).

(g) Limitation 1.f

84. Limitation 1.f recites "sequencing the amplicons to produce 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 interest*** (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence).[130]

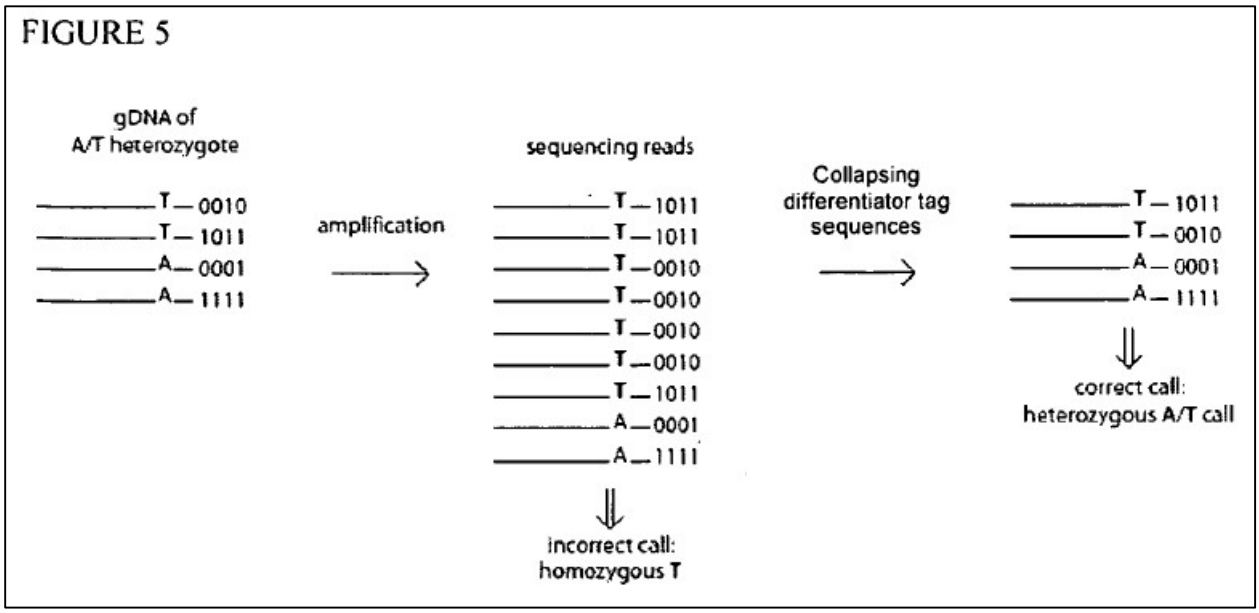
The '202 Patent Pub. (Porreca) also explains that the sequencing reads contain both differentiator and target sequences. "[T]he differentiator tag sequence permits the detection of bias based on the frequency with which ***pairs of differentiator tag and target sequences are observed in a sequencing reaction***."[131] The '202 Patent Pub. (Porreca) further explains that "***[t]he sequencing primers must be incorporated***

[130] *Id*. at [0142] (emphasis added).

[131] *Id*. at [0138] (emphasis added).

subsequently such that they sequence both the differentiator tag sequence and the target sequence."[132]

85. 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*.



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."[133]

---

[132] *Id*. at [0138] (emphasis added).

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

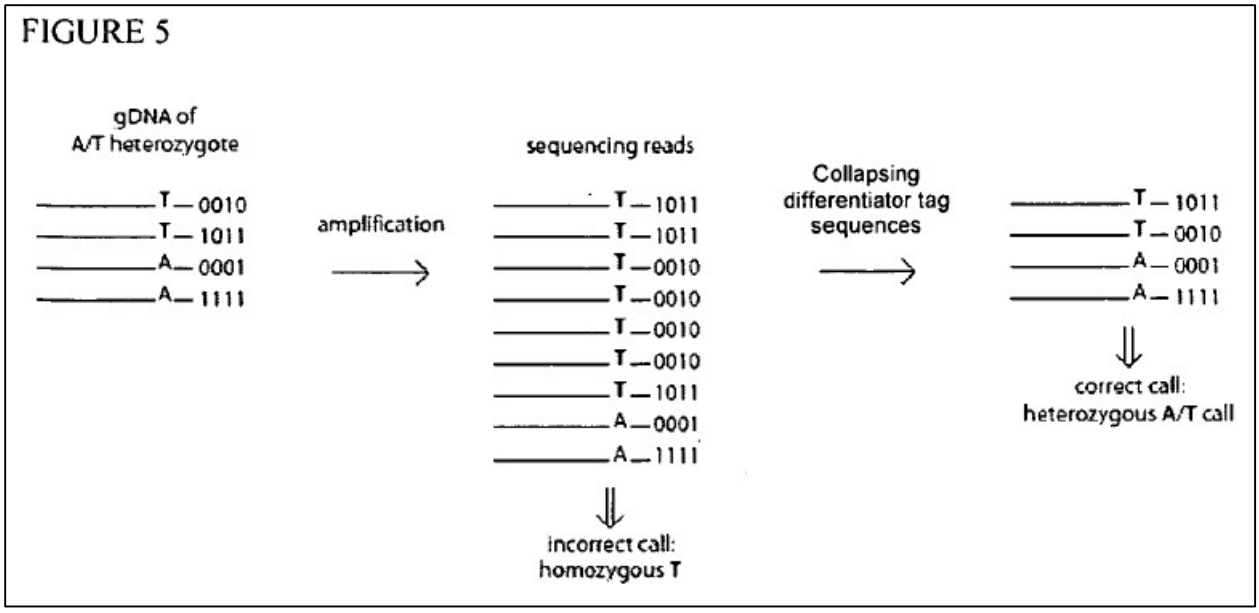
(h) Limitation 1.g

86. Limitation 1.g 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 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.

87. 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***)."[134] 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*.

---

[134] *Id*. at [0024] (emphases added).



88. 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***."[135] 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 differentiator tag sequences.

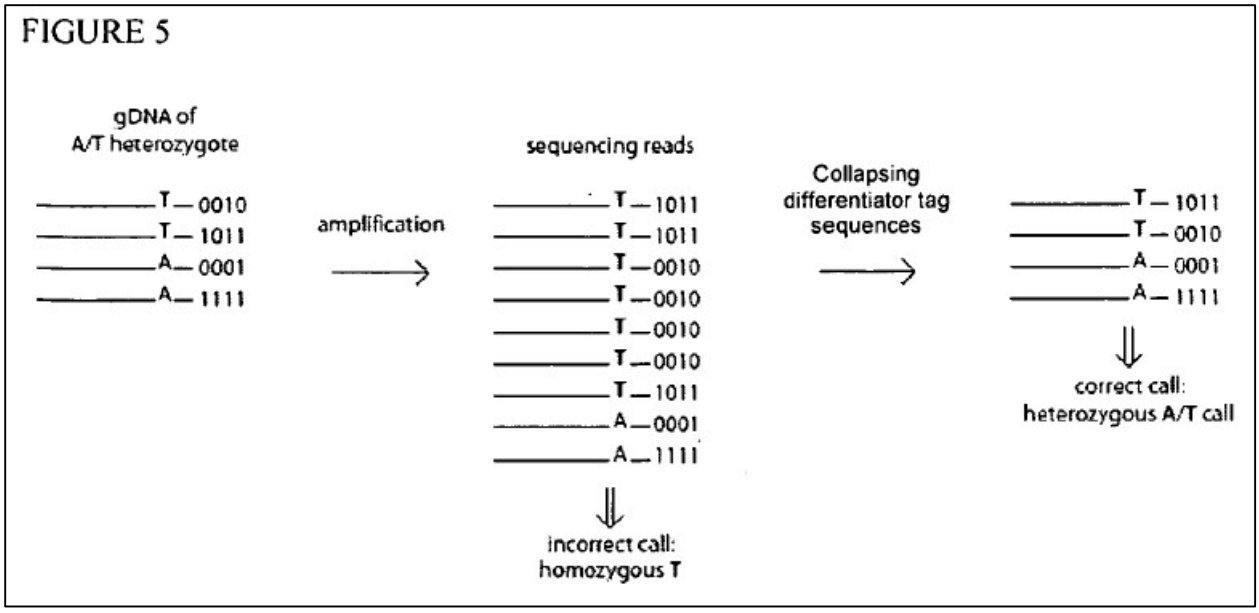
[135] *Id*. at [0148] (emphases added).

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***.”[136]

2. Claim 2 of the ’357 Patent

89. 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).

[136] *Id*. at [0298] (emphasis added).



90. 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 in that group. I note that the '357 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."[137] 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***."[138] Thus, the '202 Patent Pub. (Porreca)'s description of collapsing duplicate reads into a consensus read satisfies this claim.

3. Claim 3 of the '357 Patent

91. Claim 3 recites "[t]he method of claim 1, wherein the identifier site is sequenced with the target sequence." 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*** . . . ."[139] The '202 Patent Pub. (Porreca) also

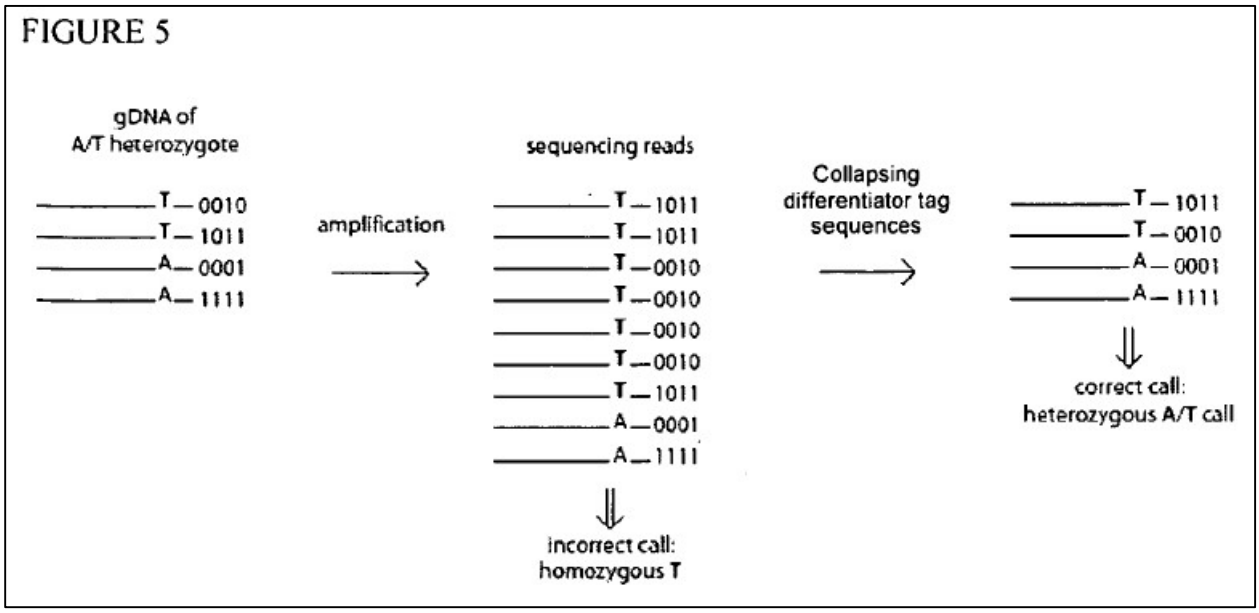
---

[137] *See* Ex. 1002, '357 Patent at 1:42-46 (emphasis added).

[138] *See* Ex. 1006, '202 Patent Pub. (Porreca) at [0298] (emphasis added).

[139] *Id*. at [0142] (emphasis added).

explains that "[i]n some embodiments, the ***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***."[140] 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***."[141] 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*.



[140] *Id*. at [0138] (emphasis added).

[141] *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 ’357 Patent

92. Claim 5 recites “[t]he method of claim 1, wherein the nucleic acid fragments comprise RNA.” 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 RNA.

> 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.”[142]

Thus, the ’202 Patent Pub. (Porreca)’s description of RNA satisfies the claim.

5. Claim 6 of the ’357 Patent

93. Claim 6 recites “[t]he method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.” 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

[142] *Id*. at [0200] (emphases added).

> as ***genomic DNA***, mitochondrial DNA, mRNA, cDNA, rRNA, miRNA, or a combination thereof.[143]

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."[144]

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

94. In my opinion, the '202 Patent Pub. (Porreca)'s disclosure of cDNA also satisfies 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."[145]

The '202 Patent Pub. (Porreca) further teaches that the differentiator tag may be incorporated at the cDNA level. "Another method to incorporate the differentiator

---

[143] *Id*. at [0200] (emphases added).

[144] *Id*. at [0147] (emphasis added).

[145] *Id*. at [0200] (emphases added).

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."***[146] Thus, the '202 Patent Pub. (Porreca)'s description of cDNA or reverse transcribing RNA into cDNA satisfies the claim.

6. Claim 7 of the '357 Patent

95. Claim 7 recites "[t]he method of claim 1, 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***."[147] 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***."[148]

---

[146] *Id*. at [0150] (emphasis added).

[147] *Id*. at [0149] (emphasis added).

[148] *Id*. at [0150] (emphases added).

96. The ’202 Patent Pub. (Porreca) also discloses “common adaptor.”[149] 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.[150]

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.

---

[149] *See* Ex. 1002, ’357 Patent at 10:5-9 (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).

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

7. Claim 8 of the ’357 Patent

97. Claim 8 recites “[t]he method of claim 1, wherein each unique identifier is 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.”[151] Thus, the ’202 Patent Pub. (Porreca)’s description of the length of the unique identifier as “up to 6” satisfies the claim.

8. Claim 9 of the ’357 Patent

98. Claim 9 recites “[t]he method of claim 1, wherein each unique identifier is 8 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

[151] *Id*. at [0141] (emphasis added).

more nucleotides in length."[152] Thus, the '202 Patent Pub. (Porreca)'s description of the length of the unique identifier as "up to 8" satisfies the claim.

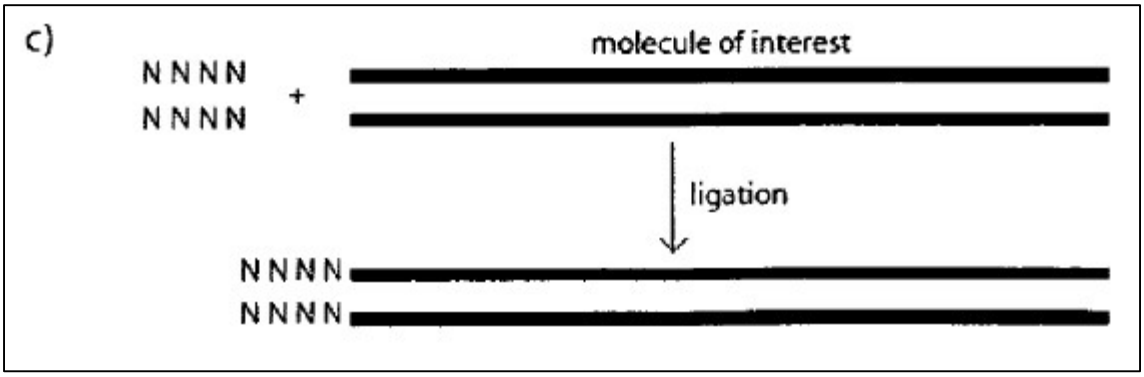
9. Claim 10 of the '357 Patent

99. Claim 10 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 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***."[153] 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*.

---

[152] *Id*. at [0141] (emphasis added).

[153] *Id*. (emphasis added).



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

B. Ground II: Claims 1-3, 5-10 Would Have Been Obvious Over the ’389 PCT Appl. (Schmitt) In View Of the ’640 Patent Pub. (McCloskey)

100. In my opinion, Claims 1-3, 5-10 of the ’357 Patent would have been obvious over the ’389 PCT Appl. (Schmitt) in view of the ’640 Patent Pub. (McCloskey). Below is a table matching the claim terms in the ’357 Patent to the corresponding disclosure in the ’389 PCT Appl. (Schmitt) and the ’640 Patent Pub. (McCloskey).

’241 patent	**’389 PCT Appl. (Schmitt)**	**’640 Patent Pub. (McCloskey)**
“identifier site”	“single molecule identifier (SMI)”	“barcode”
“indexing site”	“barcode”	“batch-stamp”

1. Independent Claim 1 of the ’357 Patent

(a) Limitation 1.p

101. 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***.”[154] 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.”[155] 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 from

---

[154] *See* Ex. 1007,’389 PCT Appl. (Schmitt) at [0085] (emphases added).

[155] *Id*. at [0098] (emphasis added).

the family of duplicates. . . . ***The improved sequence coverage arose from use of the SMI to identify PCR duplicates***."[156]

(b) Limitation 1.a

102. Limitation 1.a recites "ligating an adaptor to each of a plurality of nucleic acid fragments." 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 allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence***."[157]

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."[158] 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***."[159]

---

[156] *Id*. at [0104] (emphases added).

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

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

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

103. 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-HCI pH 7.6, 10 mM MgCl2, 5 mM DTT, 1 mM ATP."[160] Thus, the '389 PCT Appl. (Schmitt)'s description of a ligation reaction satisfies the claim.

(c) Limitation 1.b

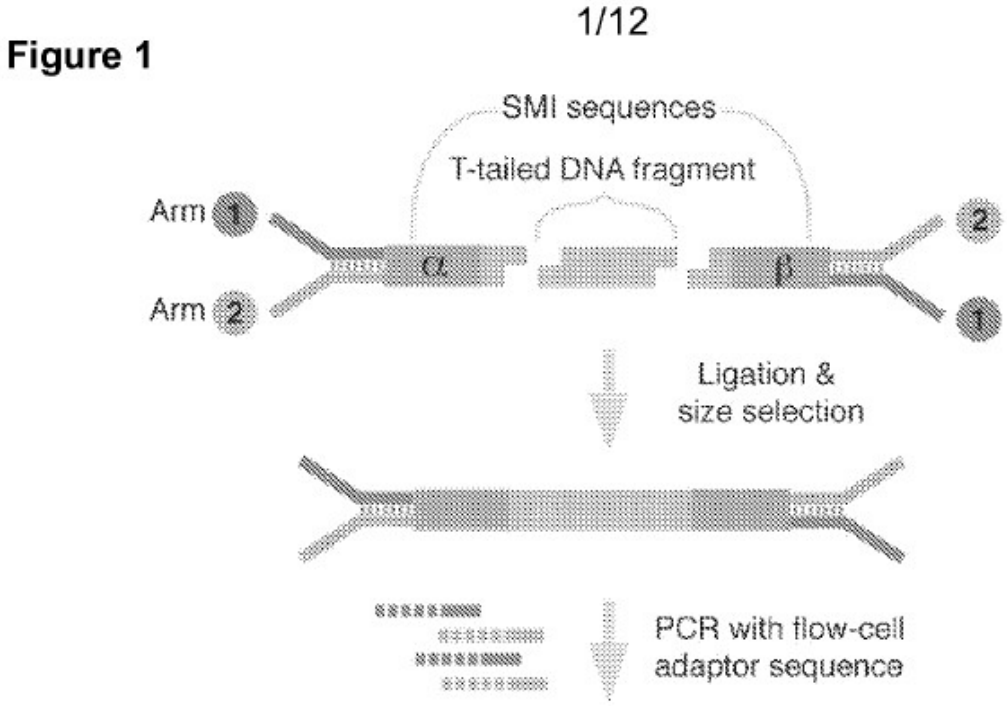
104. Limitation 1.b recites "wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides." In my opinion, the '389 PCT Appl. (Schmitt) discloses this limitation. For example, the '389 PCT Appl. (Schmitt)'s SMI adaptor molecule satisfies the claimed "identifier site." The '389 PCT Appl. (Schmitt) discloses that the SMI sequence is made up of nucleotides, including 4-8 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***

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

> ***first and/or second degenerate or semi-degenerate nucleotide n-mer sequences may be any suitable length*** to produce a sufficiently large number of unique tags to label a set of sheared DNA fragments from a segment of DNA. ***Each n-mer sequence may be between approximately 3 to 20 nucleotides in length***. Therefore, each n-mer sequence may be approximately ***4, 5, 6, 7, 8***, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides in length. In one embodiment, ***the SMI sequence is a random degenerate nucleotide n-mer sequence which is 12 nucleotides in length***. A 12 nucleotide SMI n-mer sequence that is ligated to each end of a target nucleic acid molecule, as described in the Example below, results in generation of up to 4^{24} (i.e., 2.8×10^{14}) distinct tag sequences.[161]

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



[161] *Id*. at [0026] (emphases added).

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."[162] The '389 PCT Appl. (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."[163]

106. 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***."[164]

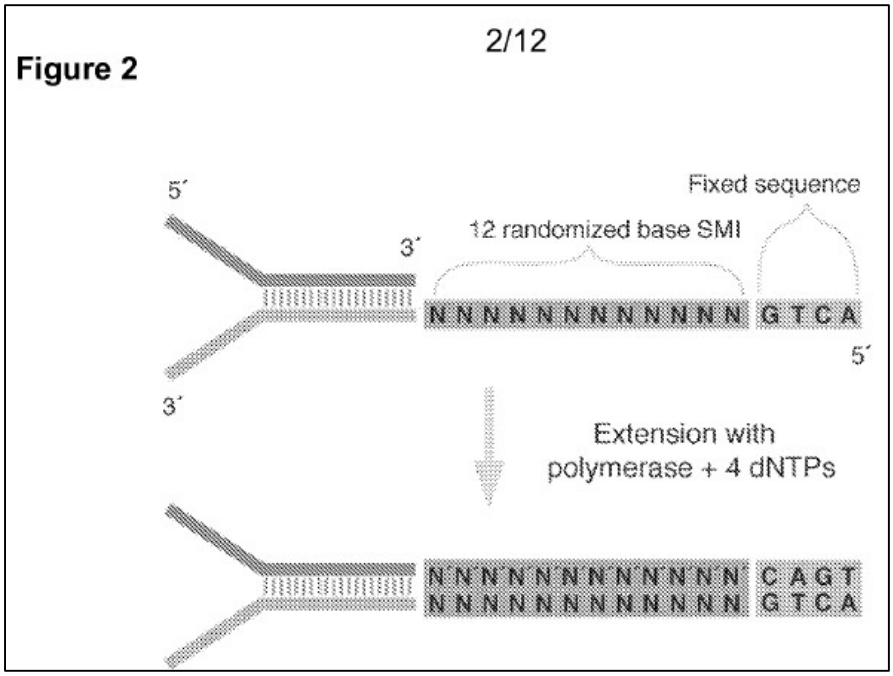
107. Figure 2 of the '389 PCT Appl. (Schmitt) shows an exemplary 12 nucleotide SMI from Example 1, as reproduced *infra*.

---

[162] *Id*. at [0012] (emphasis added).

[163] *Id*. at [0085] (emphasis added).

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



(d) Limitation 1.c

108. In my opinion, it would have been obvious to include the ’640 Patent Pub. (McCloskey)’s batch-stamp (the claimed indexing site) in the ’389 PCT Appl. (Schmitt)’s adaptor. This combination satisfies Limitation 1.c.

(i) One of ordinary skill in the art would have been motivated to add an indexing site to the ’389 PCT Appl. (Schmitt)’s adaptor

109. In my opinion, one of ordinary skill in the art would have been motivated to include the claimed indexing site in the ’389 PCT Appl. (Schmitt)’s adaptor, described *supra* at Limitation 1.a. The ’389 PCT Appl. (Schmitt)’s Example 1 discloses that, after the DNA capture step, the adaptor appended products were amplified with PCR primers that included a “fixed multiplexing barcode

sequence."[165] One of ordinary skill in the art would have known that a multiplexing barcode sequence allows for the identification of the sample from which a particular sequence read originated. Multiplex analysis of multiple samples provides several advantages. For instance, combining multiple samples together allows screening to be done in a more efficient way, as sequencing reads from multiple samples can be obtained in a single run. In order to effectively conduct this type of analysis, all sequencing reads must be traced to their sample of origin. This is performed using barcodes unique for every sample.

110. The '389 PCT Appl. (Schmitt)'s barcode sequence, however, is appended to the target DNA sequences after several steps of the disclosed method had been performed. Library preparation, pre-capture amplification, and DNA capture steps all occur with the barcode was appended.[166] One of ordinary skill in the art would have known that it would be beneficial to append a barcode to the target sequences earlier in the process to reduce the chance of errors that may affect the analysis, such as mixing different samples together prior to labeling. One of ordinary skill in the art thus would have been motivated to include the multiplexing barcodes to the adaptor described in Limitation 1.a, which occurs before the first amplification and DNA capture steps.

---

[165] *Id.* at [0095].

[166] *Id.* at [0092[-[0095].

(ii) the ’640 Patent Pub. (McCloskey) discloses the claimed indexing site

111. In my opinion, the ’640 Patent Pub. (McCloskey) discloses “an indexing site unique to a subset of the adaptors.” The ’640 Patent Pub. (McCloskey) discloses a hairpin linker adaptor embodiment and an oligonucleotide adaptor embodiment. Both of these embodiments include a “batch-stamp,” which satisfies the claimed “indexing site.” For example, 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.”[167]

112. The hairpin linker embodiment includes a batch-stamp. “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.” Figure 3 is reproduced *infra*.

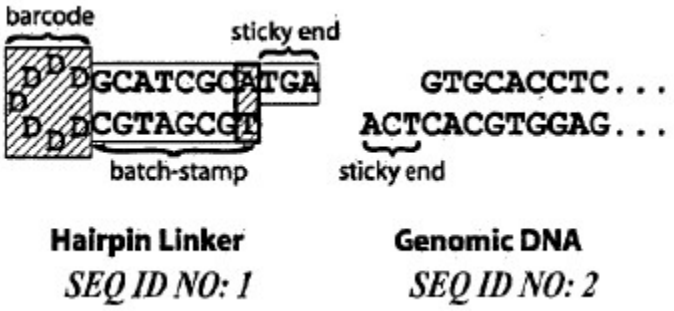


Fig.3.

[167] *See* Ex. 1005, ’640 Patent Pub. (McCloskey) at [0027].

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-GATGCDDDDDDD GCATCGCTTGA 3’, SEQ ID NO:1) with variations in the non-random nucleotides for batch-stamps . . . .”[168]

113. The 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.”[169]

(iii) One of ordinary skill in the art would have been motivated to combine the ’389 PCT Appl. (Schmitt)’s adaptor with the ’640 Patent Pub. (McCloskey)’s indexing site

114. One of ordinary skill in the art would have been motivated to combine the ’389 PCT Appl. (Schmitt) with the ’640 Patent Pub. (McCloskey)’s batch-stamp. Both references are in the same field of endeavor and contain disclosures that solve the same problem. The’389 PCT Appl. (Schmitt) and the ’640 Patent Pub. (McCloskey) both utilize identifier and indexing sites to identify PCR duplicates and to match sequencing reads to the originating sample. For instance, the ’640 Patent

---

[168] *Id.* at [0095].

[169] *Id.* at [0013]; *see also* Examples 1-5.

Pub. (McCloskey) discloses the use of "barcodes" to distinguish "valid sequences" from "redundant sequences" (*i.e.*, duplicate sequences),[170] as well as the use of a "batch-stamp" to identify which sample each sequencing read originated from.[171] The '389 PCT Appl. (Schmitt) has similar teachings.[172]

115. The '389 PCT Appl. (Schmitt) also contains a citation to McCloskey (2007)[173] McCloskey (2007) and the '640 Patent Pub. (McCloskey) share some disclosures, such as an oligonucleotide adaptor that includes an identifier and indexing site.[174] However, the '640 Patent Pub. (McCloskey) contains a more substantial disclosure and material that is not reported in McCloskey (2007), such as the hairpin linker adaptor embodiment, sequencing data from Examples 1-6, and discussion about data analysis techniques. In my opinion, one of ordinary skill in the art would have been aware of the '640 Patent Pub. (McCloskey), particularly in light of the '389 PCT Appl. (Schmitt)'s explicit citation to McCloskey (2007). One of ordinary skill in the art would have been aware of the '640 Patent Pub.

---

[170] *Id.* at [0005].

[171] *See supra* § X.B.1(d)(ii).

[172] *See supra* §§ X.B.1(c), X.B.1(d)(i).

[173] *See* Ex. 1015, McCloskey *et al.*, *Encoding PCR products with batch-stamps and barcodes*, BIOCHEMICAL GENETICS 45:761-767 (2007) ("McCloskey (2007)"). *See* also Ex. 1007, '389 PCT Appl. (Schmitt) at [0006], citing reference #35, which is McCloskey (2007).

[174] *Compare* Ex. 1015, McCloskey (2007)'s Figure 1, *with* Ex. 1005, '640 Patent Pub. (McCloskey)'s Figure 1.

(McCloskey) that is associated with McCloskey (2007). One of ordinary skill in the art further would have been motivated to select the ’640 Patent Pub. (McCloskey) as a reference to combine with the ’389 PCT Appl. (Schmitt), as the ’640 Patent Pub. (McCloskey)’s batch-stamp allows the ’389 PCT Appl. (Schmitt)’s adaptor to be used in multiplex sequencing reactions.

116. One of ordinary skill in the art also would have had at least a reasonable expectation of success from pursuing the combination. It was well-known in the art that indexing sites could be included in adaptors. One of ordinary skill in the art thus would have known how to include the ’640 Patent Pub. (McCloskey)’s batch-stamp in the ’389 PCT Appl. (Schmitt)’s adaptor. The ’389 PCT Appl. (Schmitt) teaches that adaptors may contain different types of sequences. For example, the ’389 PCT Appl. (Schmitt)’s adaptor can include an “SMI adaptor molecule,” “PCR primer binding sites,” “sequencing primer binding sites,” or a “multiplexing barcode sequence.”[175] The ’389 PCT Appl. (Schmitt) further teaches that there may be multiple sequences on a single adaptor. “Optionally, the SMI adaptor molecule further includes at least two PCR primer binding sites, at least two sequencing primer binding sites, or both.”[176] One of ordinary skill in the art thus would have known

[175] *See* Ex. 1007, ’389 PCT Appl. (Schmitt) at [0010], [0022], [0095].

[176] *Id.* at [0022].

how to include different types of sequences into an adaptor, such as an SMI sequence and the '640 Patent Pub. (McCloskey)'s batch-stamp.

(e) Limitation 1.d

117. Limitation 1.d recites "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***."[177] 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***."[178]

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)."[179]

---

[177] *Id*. at [0022] (emphasis added).

[178] *Id*. at [0022] (emphasis added).

[179] *Id*. at [0031] (emphasis added).

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

4/12

Figure 4

A

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

B

```
5'   1-TAAC----------------------TCCG-2   3'   (top strand)

5'   1-CGGA----------------------GTTA-2   3'   (bottom strand)
```

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.”[180] 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***.”[181] Thus,

[180] *Id*. at [0015] (emphases added).

[181] *Id*. at Claim 5 (emphasis added).

the ’389 PCT Appl. (Schmitt)’s description of an adaptor that includes a primer binding site satisfies the claim.

(f) Limitation 1.e

119. Limitation 1.e recites “amplifying the adaptor-ligated fragments into amplicons.” 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.[182]

120. 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), resulting in a set of uniquely labeled, amplified SMI-target nucleic acid products.[183]

---

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

[183] *Id*. at [0074] (emphasis added).

121. 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***.[184]

122. 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***.[185]

The ’389 PCT Appl. (Schmitt) further explains that “***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.”[186]

---

[184] *Id*. at [0010] (emphases added).

[185] *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].

[186] *Id*. at [0093] (emphasis added).

(g) Limitation 1.f

123. Limitation 1.f 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***.[187]

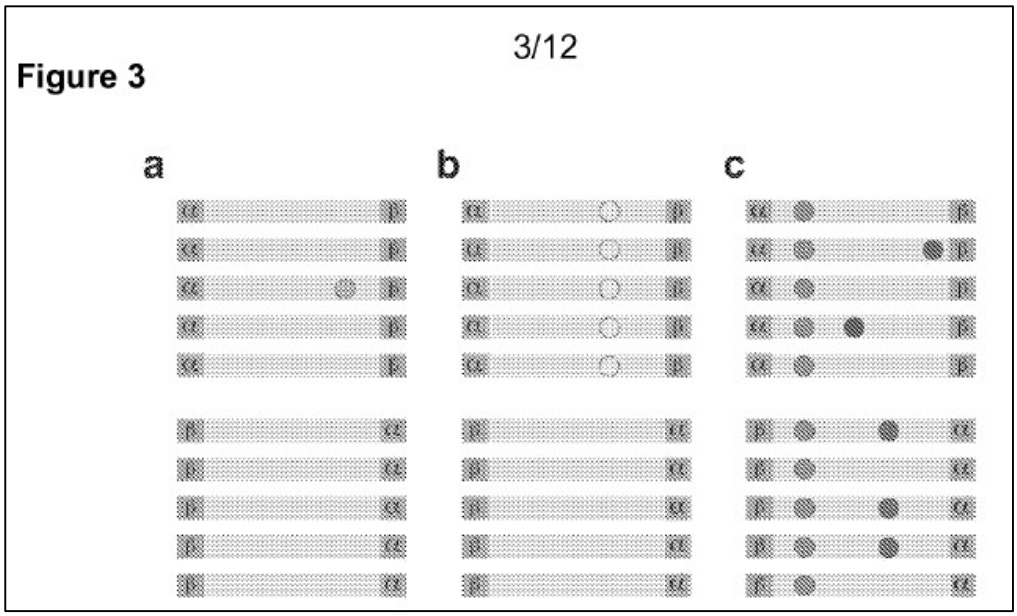
124. The '389 PCT Appl. (Schmitt) explains that the amplicons can be sequenced with any known methods in the art.

> 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

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

sequencing platform, Helicos sequencing platform, and nanopore sequencing technology.[188]

125. 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*.

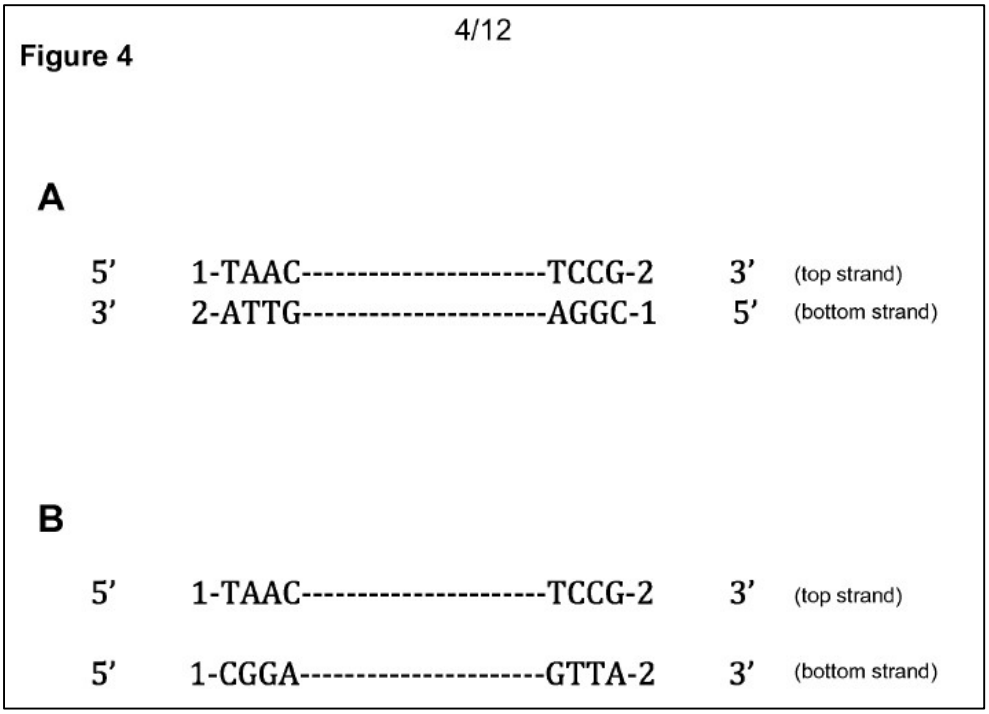


The ’389 PCT Appl. (Schmitt) explains that “(a-c) shows ***sequence reads (brown) sharing a unique set of SMI tags*** are grouped into paired families with members having strand identifiers in either the αβ or βα orientation.”[189] Figure 4 of the ’389 PCT Appl. (Schmitt) also shows an example of sequencing reads of both the identifier and target sequences, as reproduced *infra*.

---

[188] *Id*. at [0074] (emphasis added).

[189] *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."[190]

[190] *Id*. at [0014] (emphasis added).

126. With respect to Example 1, the '389 PCT Appl. (Schmitt) explains that SMI-ligated DNA was PCR amplified and sequenced.[191] "[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***."[192] 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.[193]

(h) Limitation 1.g

127. Limitation 1.g 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

---

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

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

[193] *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***."[194] 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***.[195]

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***.[196]

128. 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***

---

[194] *Id*. at [0085] (emphasis added).

[195] *Id*. at [0098] (emphases added).

[196] *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***."[197] 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***."[198] 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 '357 Patent

129. 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***

[197] *Id*. at [0099] (emphases added).

[198] *Id*. at [0096] (emphasis added).

a consensus read."[199] 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.[200]

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 '357 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***."[201] Thus, the '389 PCT Appl. (Schmitt)'s description of collapsing duplicate reads into a consensus read satisfies this claim.

3. Claim 3 of the '357 Patent

130. 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

---

[199] *Id*. at [0096] (emphasis added).

[200] *Id*. (emphases added).

[201] *See* Ex. 1002, '357 Patent at 1:41-45 (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***.[202]

131. 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 αβ or βα orientation."[203] 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---***."[204] 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***."[205] The reads

---

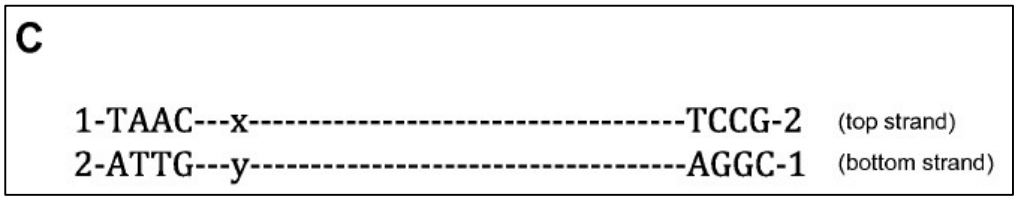
[202] *See* Ex. 1007,'389 PCT Appl. (Schmitt) at [0010] (emphasis added).

[203] *Id*. at [0014] (emphasis added).

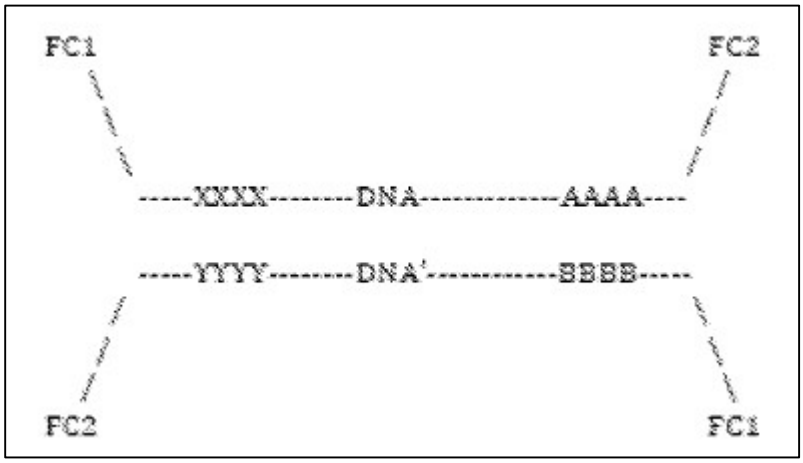
[204] *Id*. at [0015] (emphasis added).

[205] *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 II[206] 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).[207]

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

[206] *Id*. at [0042] (Product II).

[207] *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."[208] 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***."[209] 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 '357 Patent

133. Claim 5 recites "[t]he method of claim 1, wherein the nucleic acid fragments comprise RNA." 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

---

[208] *Id*. at [0096] (emphases added).

[209] *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.[210]

Thus, the '389 PCT Appl. (Schmitt)'s description RNA satisfies the claim.

5. Claim 6 of the '357 Patent

134. Claim 6 recites "[t]he method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments." 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."[211] The '389 PCT Appl. (Schmitt) also explains that "[***g]enomic DNA was isolated*** from a derivative of *Saccharomyces cerevisiae* strain SC288 by standard methods."[212] Thus, the '389 PCT Appl. (Schmitt)'s description of a genomic DNA satisfies the claim.

135. In my opinion, the '389 PCT Appl. (Schmitt)'s disclosure of cDNA also satisfies 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 to sequencing of DNA molecules, ***but are also applicable to RNA***

---

[210] *Id*. at [0072] (emphasis added).

[211] *Id*. at [0089] (emphasis added).

[212] *Id*. at [0114] (emphasis added).

> ***sequencing applications*** such as for sequencing of double-stranded RNA viruses.[213]

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."[214] 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.

6. Claim 7 of the '357 Patent

136. Claim 7 recites "[t]he method of claim 1, 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

---

[213] *Id*. at [0072] (emphasis added).

[214] *Id*. (emphasis added).

> sequences as tails to the PCR primers, or by ligation of the needed sequences.[215]

137. 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 of the sequencing primers***, which generate the complementary sequence during read 2.[216]

138. 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.[217]
>
> 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.[218]

---

[215] *Id*. at [0031] (emphasis added).

[216] *Id*. at [0015] (emphases added).

[217] *Id*. at Claim 5 (emphases added).

[218] *Id*. at Claim 28 (emphasis added).

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.

7. Claim 8 of the '357 Patent

139. Claim 8 recites "[t]he method of claim 1, wherein each unique identifier is 6 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 from 3-20 nucleotides in length and may be approximately 6 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"[219]

[219] *Id*. at [0026] (emphases added).

8. Claim 9 of the '357 Patent

140. Claim 9 recites "[t]he method of claim 1, wherein each unique identifier is 8 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 from 3-20 nucleotides in length and may be approximately 8 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"[220]

9. Claim 10 of the '357 Patent

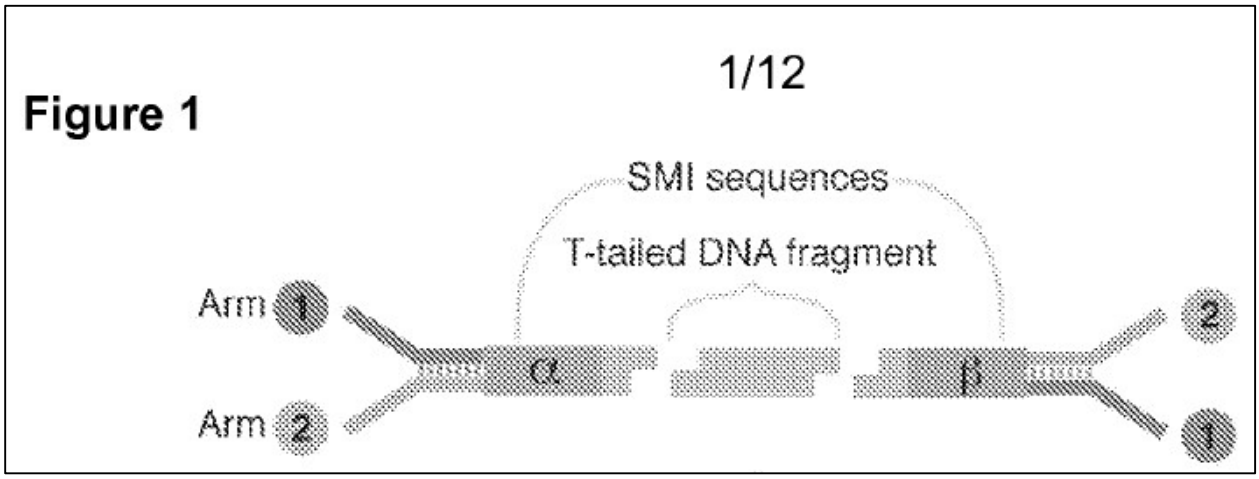
141. Claim 10 recites "[t]he method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment." In my opinion, the '389 PCT Appl. (Schmitt) discloses this limitation. I understand that during prosecution of a related patent, the examiner found that the term "junction" is considered as any region of the adaptor.[221] The '389 PCT Appl. (Schmitt)'s Figure

---

[220] *Id*. at [0026] (emphases added).

[221] *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 adapter***.") (emphasis added).

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*.



142. The '389 PCT Appl. (Schmitt) explains that the unique identifier is between the adapter 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 sequence adjacent to flow-cell sequence 1 and the β SMI sequence adjacent to flow cell sequence 2***.[222]

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

[222] *See* Ex. 1007,'389 PCT Appl. (Schmitt) at [0012] (emphases added).

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.

C. Grounds III and IV: Claim 4 of the '357 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 '640 Patent Pub. (McCloskey) in further view of Illumina (2008)

143. 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 produces 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 sites

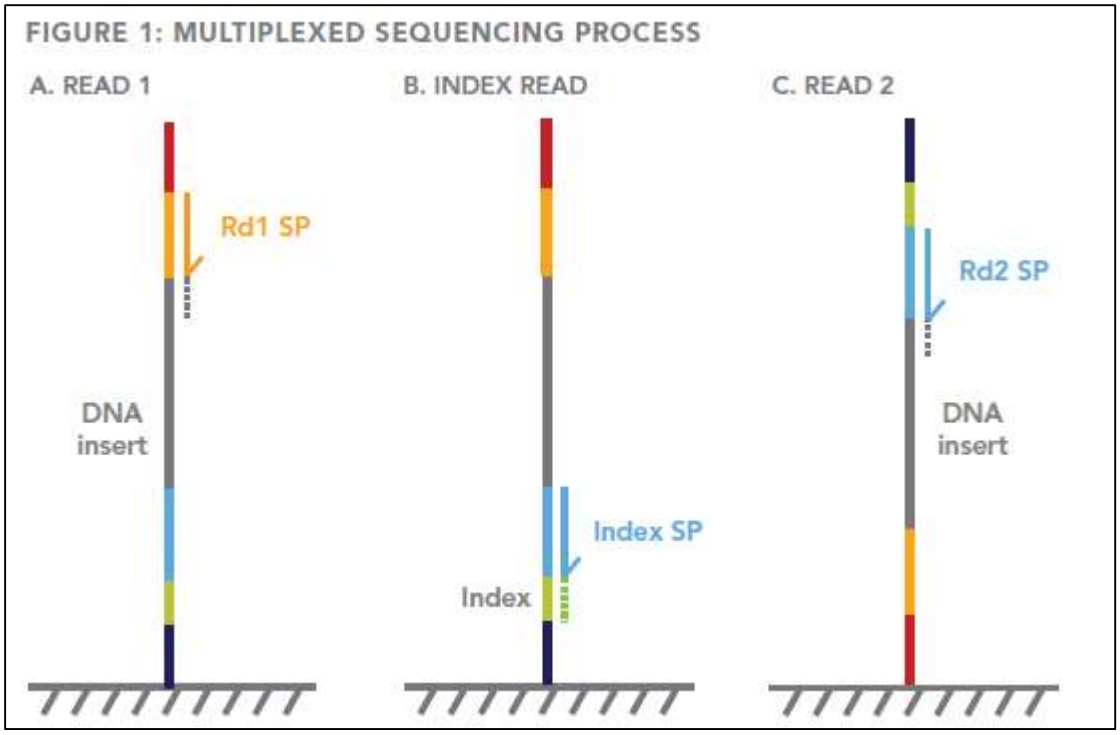
separately, particularly with the Illumina platform that produces short sequence reads.

144. 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 '357 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***."[223] 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*.[224]

---

[223] *See* Ex. 1010, Metzker (2010) at 34 (emphasis added).

[224] *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."[225] 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 have included an identifier sequence in the same location as the index sequence.

[225] *Id*. at 1.

145. 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)."[226] 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 . . . ."[227] 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."[228] 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."[229]

---

[226] *See* Ex. 1006, '202 Patent Pub. (Porreca) at [0295].

[227] *See* Ex. 1007, '389 PCT Appl. (Schmitt) at [0074].

[228] *Id.* at [0086] (emphasis added).

[229] *Id*. at [0083].

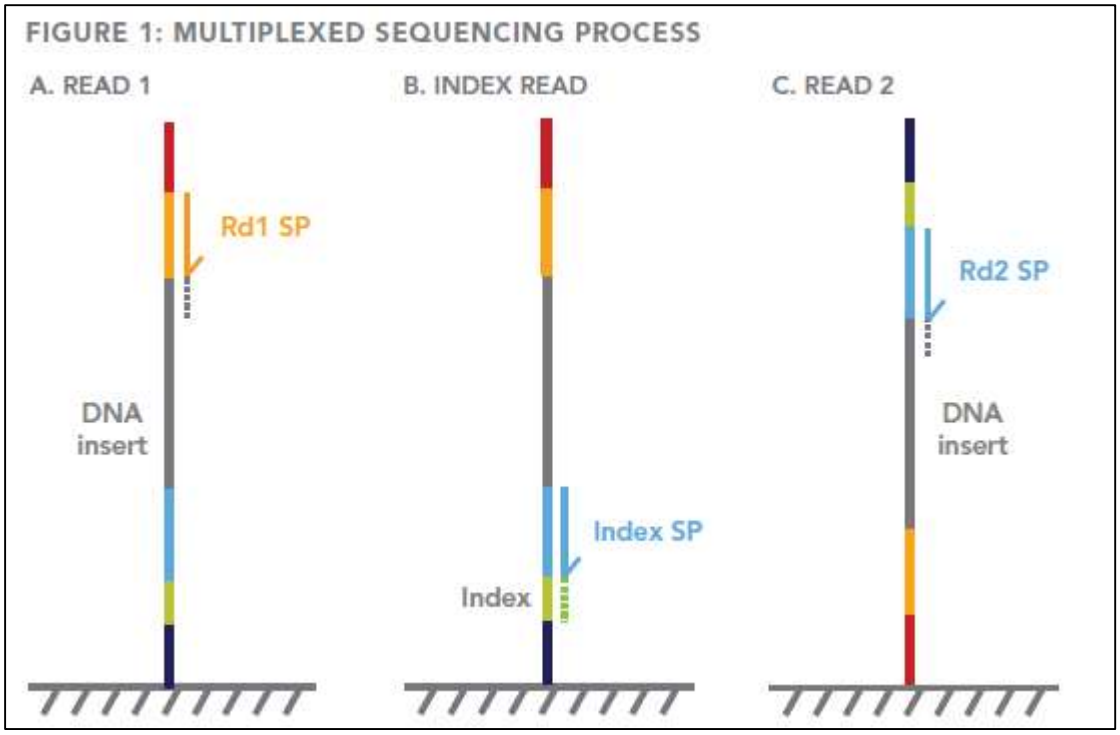
D. Grounds V and VI: Claim 7 of the ’357 Patent would have been obvious over the ’202 Patent Pub. (Porreca) in view of Illumina (2008) and over ’389 PCT Appl. (Schmitt) in view of ’640 Patent Pub. (McCloskey) in further view of Illumina (2008)

146. As I discuss *supra* in Grounds I 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.

147. 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 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*.[230]

---

[230] *See* Ex. 1008, Illumina (2008) at 1.



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*.

148. 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)***."[231] One of ordinary skill in the art thus would have known how

[231] *Id*. at Figure 1 legend (emphases added).

to successfully include the claimed universal target sequencing primer binding site into the ’202 Patent Pub. (Porreca) and the ’389 PCT Appl. (Schmitt) adaptors.

149. 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.”[232] The Illumina platform therefore provides a universal target sequencing primer binding site that was used in sequencing in the ’202 Patent Pub. (Porreca) and the ’389 PCT Appl. (Schmitt) methods.

150. 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

[232] *See* Ex. 1010, Metzker (2010) at 32 (emphases added).

sequencing step. "***Perform Sequencing of Library According to Manufacturer's Directions (e.g. Illumina***, ABI, etc)."[233] The '389 PCT Appl. (Schmitt) also provides a similar disclosure. "***These products are then sequenced using any suitable method known in the art inclu***ding, but not limited to, ***the Illumina sequencing platform***, ..."[234] 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 lllumina DNA sequencing***."[235]

---

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

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

[235] *Id*. at [0086] (emphasis added).

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.