

UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE PATENT TRIAL AND APPEAL BOARD

FORESIGHT DIAGNOSTICS INC.,
Petitioner,

v.

PERSONALIS, INC.,
Patent Owner.

IPR2024-00170
U.S. Patent No. 11,584,968

**DECLARATION OF JOHN QUACKENBUSH, PH.D. IN SUPPORT OF
PETITIONER FOR *INTER PARTES* REVIEW**

Mail Stop: PATENT BOARD
Patent Trial and Appeal Board
U.S. Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

TABLE OF CONTENTS

	<u>Page</u>
I. Introduction	1
II. Legal Standards	7
A. Claim Construction	7
B. Anticipation.....	8
C. Obviousness	8
III. The ‘968 Patent.....	12
IV. Person of Ordinary Skill in the Art	12
V. Overview of a Person of Ordinary Skill in the Art’s Knowledge	14
A. Overview of Molecular Biology and Genetics	14
1. The Central Dogma of Molecular Biology.....	14
2. Genetics.....	18
B. Genetic Mutations	20
C. Nucleic Acid Sequencing Technologies	22
1. Sequencing Techniques	22
2. Sequencing Depth and Read Counts	26
3. Amplification and Enrichment	30
D. Personalized Cancer Testing and Monitoring.....	38
E. Related Applications of Next Generation Sequencing	45
F. Providing Actionable Information to Healthcare Providers.....	48
VI. Prior Art References of the Grounds.....	51

TABLE OF CONTENTS
Cont'd

	<u>Page</u>
A. Chan (Ex. 1008)	51
B. Okosun-II (Ex. 1102)	51
C. Ley (Ex. 1012)	52
VII. Claim Construction.....	52
A. “A method of analyzing a biological sample obtained from a subject being screened for a cancer” – Claim 1	52
B. “mosaic variants” – Claim 1	53
C. “capture probe” – Claim 13	55
VIII. Prior Art Anticipated or Would Have Rendered Obvious Claims 1-13 of the ‘968 Patent.....	57
A. Ground 1: Chan Anticipates Claims 1-4, 7, and 9-12.....	57
1. Claim 1	57
2. Claim 2	68
3. Claim 3	69
4. Claim 4	70
5. Claim 7	71
6. Claim 9	71
7. Claim 10	72
8. Claim 11	72
9. Claim 12	73

TABLE OF CONTENTS
Cont'd

	<u>Page</u>
B. Ground 2: Claims 1-13 would have been obvious over Chan in view of the knowledge of a POSA.	74
1. Claim 2.....	74
2. Claims 5 and 6	75
3. Claim 8.....	76
4. Claim 13.....	77
C. Ground 3: Claims 1-13 Would Have Been Obvious Over Ley in view of Knowledge of a POSA	78
1. Overview of Ley	78
2. Adaptations of Ley’s Approach to Other Cancers.....	80
3. Claim 1.....	84
4. Claim 2.....	98
5. Claim 3.....	104
6. Claim 4.....	104
7. Claims 5 and 6	105
8. Claim 8.....	106
9. Claim 9.....	107
10. Claim 11.....	107
11. Claim 12.....	108
12. Claims 7, 10, and 13	109

TABLE OF CONTENTS
Cont'd

	<u>Page</u>
D. Ground 4: Claims 1, 8-9, and 11-13 Are Anticipated By Okosun-II	110
1. Claim 1	110
2. Claim 8	122
3. Claim 9	123
4. Claim 11	123
5. Claim 12	124
6. Claim 13	125
E. Ground 5: Claims 1-13 Would Have Been Obvious Over Okosun-II in view of Knowledge of a POSA	127
1. Claim 1	128
2. Claim 2	129
3. Claim 3	133
4. Claim 4	133
5. Claims 5 and 6	134
6. Claims 7, 10, and 13	135
IX. Secondary Considerations of Nonobviousness	138

Exhibits Cited

EX1001	U.S. Patent No. 10,584,968 issued to West
EX1002	Rebecca J. Leary <i>et al.</i> , “Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing,” <i>Science Translational Medicine</i> (2010) (Leary)
EX1003	Quake <i>et al.</i> , WO 2011/057061 (Quake)
EX1005	Andreas Gnirke <i>et al.</i> , “Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing,” <i>Nature Biotechnology</i> (2009) (Gnirke)
EX1006	Florian Mertes <i>et al.</i> , “Targeted enrichment of genomic DNA regions for next-generation sequencing, Briefings in Functional Genomics,” <i>Briefings in Functional Genomics</i> (2011) (Mertes)
EX1007	Lira Mamanova <i>et al.</i> , “Target-enrichment strategies for next-generation sequencing,” <i>Nature</i> (2010) (Mamanova)
EX1008	Chan et al., “Cancer Genome Scanning in Plasma: Detection of Tumor-Associated Copy Number Aberrations, Single-Nucleotide Variants, and Tumoral Heterogeneity by Massively Parallel Sequencing,” <i>Clinical Chemistry</i> , January 1, 2013
EX1010	<i>Personalis v. Foresight</i> , First Amended Complaint, D. Colo., Case No. 1:22-v-01913, Dkt. 9, filed August, 17, 2022 (“FAC”)
EX1012	Timothy J. Ley <i>et al.</i> , “DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome,” <i>Nature</i> (2008) (Ley)
EX1013	Jay Shendure & Hanlee Ji, “Next-generation DNA sequencing,” <i>Nature Biotechnology</i> (2008) (Shendure)
EX1014	Jeffrey S. Ross and Maureen Cronin, “Whole Cancer Genome Sequencing by Next-Generation Methods,” <i>American Journal of Clinical Pathology</i> (2011) (Ross)
EX1015	David J. McBride, et al., “Use of Cancer-Specific Genomic Rearrangements to Quantify Disease Burden in Plasma from

	Patients with Solid Tumors,” <i>Genes, Chromosomes & Cancer</i> (2010) (McBride)
EX1023	<i>Curriculum Vitae</i> of John Quackenbush
EX1024	https://www.garvan.org.au/research/kinghorn-centre-for-clinical-genomics/learn-about-genomics/for-gp/genetics-refresher-1/types-of-variants
EX1027	Hugo Lam, et al., “Performance comparison of whole-genome sequencing platforms,” <i>Nature Biotechnology</i> (2012) (Lam)
EX1028	Illumina, Technical Note: Sequencing
EX1029	Illumina, Specification Sheet: Illumina Sequencing
EX1031	Matthew Meyerson, et al., “Advances in understanding cancer genomes through second-generation sequencing,” <i>Nature Reviews</i> (2010) (Meyerson)
EX1040	D. Sims <i>et al.</i> , “Sequencing depth and coverage: key considerations in genomic analyses,” <i>Nature Reviews: Genetics</i> (2014) (Sims)
EX1042	Ley Supplemental Information
EX1043	A. Newman <i>et al.</i> , “Integrated digital error suppression for improved detection of circulating tumor DNA,” <i>Nature Biotechnology</i> (2016) (Newman)
EX1044	A. Newman <i>et al.</i> , “An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage,” <i>Nature Medicine</i> (2014) (Newman 2014)
EX1070	S. Dawson, <i>et al.</i> , “Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer,” <i>New England Journal of Medicine</i> (2013) (Dawson)
EX1077	D. Bentley, et al., “Accurate whole human genome sequencing using reversible terminator chemistry,” <i>Nature</i> (2008) (Bentley)

EX1101	J. Okosun, <i>et al.</i> , “Whole Genome Sequencing in Sequential Biopsies Reveals the Genetic Evolution of Follicular Lymphoma,” Blood (2012) 120 (21) 145 (Okosun I)
EX1102	J. Okosun, <i>et al.</i> , “Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma,” Nature America (2014) (Okosun II)
EX1103	Michael A. Chapman, <i>et al.</i> , “Initial genome sequencing and analysis of multiple myeloma, Nature (2011)
EX1104	William Lee, <i>et al.</i> , “The mutation spectrum revealed by paired genome sequences from a lung cancer patient,” 27 May 2010 (Lee 2010)
EX1105	Li Ding <i>et al.</i> , “Genome remodelling in a basal-like breast cancer metastasis and xenograft,”, Nature (2010) (Ding 2010)
EX1106	Yan-Fang Guan, <i>et al.</i> , “Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer,” Chinese J Cancer, 2012 Oct. 31 (10):463-470 (Guan)
EX1107	<i>Personalis v. Foresight</i> , First Amended Complaint, D. Colo., Case No. 1:23-cv-01623, Dkt. 21, filed August, 31, 2023
EX1108	Prosecution File History of Appl. 17/494,513 (Pat. No. 11,584,968)
EX1109	QIAamp® DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook, Feb. 2003 (Qiagen Protocol)
EX1110	Fluidigm, Specification Sheet for Access Array™ System, October 2012
EX1112	U.S. Publication No. 2022/0195530, issued to Diehn <i>et al.</i> , (Diehn-530)
EX1113	Supplementary Information (from Okosun II)
EX1114	Furneaux Declaration (from IPR2023-00224, -00317)

EX1115	West Declaration (from IPR2023-00224, -00317)
EX1116	Patent Owner Response from IPR2023-00224
EX1117	Dawson Supplementary Materials
EX1118	Christopher T. Saunders, <i>et al.</i> , “Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs, Bioinformatics, May 10, 2012
EX1119	Ex. F. to First Amended Complaint (Case 2)

I. Introduction

1. My name is John Quackenbush. I have been retained by Foresight Diagnostics Inc. (“Petitioner”) to provide my independent evaluation and opinions regarding U.S Patent No. 11,584,968 (the ‘968 patent) and related prior art.

2. For this service, I am paid my standard hourly consulting fee, and I am reimbursed for actual expenses. My compensation in no way depends on the outcome of this matter or the substance of this declaration.

3. In my testimony, I have relied upon my technical background, skills, knowledge, training, education and experience in the relevant art, and my examination of the materials utilized in preparing this declaration.

4. In providing this testimony, I have relied on the materials cited herein, as well as my knowledge and experience working in the field before the priority date of the ‘968 patent, from the perspective of a person of ordinary skill in the art as of the priority date.

5. I am a Henry Pickering Walcott Professor of Computational Biology and Bioinformatics and Chair of the Department of Biostatistics at the Harvard T.H. Chan School of Public Health. I am also Founding Director and co-Director of the MS Program in Computational Biology and Quantitative Genetics and Director of the Big Data to Knowledge Training Program, co-Director of the Cancer Genomics Training Program, and the Faculty Director for the Harvard Quantitative Biomedical

Research Center. In addition, I hold appointments as Professor of Biostatistics and Computational Biology at the Dana-Farber Cancer Institute and at the Channing Division of Network Medicine of Brigham and Women's Hospital.

6. In addition to my academic appointments, I co-founded Genospace LLC in 2012. Genospace produces software solutions to operationalize precision medicine through the integration of clinical and genomic data and its presentation to clinical and research users in an intuitive format. Genospace was purchased by Hospital Corporation of America in January of 2017.

7. In addition to my academic and research responsibilities, over the years I have taken significant regional and national research leadership responsibilities. I have served on the Scientific Advisory Board of the St. Jude's Children's Research Hospital and the Board of Scientific Counselors of the Environmental Protection Agency. I have also served on numerous scientific review panels for the National Institutes of Health, the National Science Foundation, the Department of Energy, the National Aeronautics and Space Administration, the Wellcome Trust, Genome British Columbia, Genome Canada, the National Sciences and Engineering Council of Canada, and the European Union. And I have served on numerous study panels for the National Research Council, the National Academy of Sciences, and the Institute of Medicine.

8. I have presented more than 350 talks, courses, and workshops since joining the Harvard School of Public Health and Dana-Farber in 2005. I have also taught numerous courses and workshops on the analysis of genomic data and served as an organizer of the EMBO course on DNA microarray analysis for more than 5 years.

9. Prior to joining the faculty at Dana-Farber and Harvard, I was a Professor of Chemical Engineering at the University of Maryland from 2003 – 2005; an Investigator with The Institute for Genomic Research from 2002 – 2006; a Professor in the Department of Biochemistry at The George Washington University from 2000 – 2005; a Lecturer in The Department of Biostatistics at The Johns Hopkins University from 1998 – 2005; an Associate Investigator at The Institute for Genomic Research from 1997 – 2001; and an Assistant Investigator at The Institute for Genomic Research in 1997.

10. I received my Ph.D. in Theoretical Particle Physics at the University of California, Los Angeles in 1990. I also hold a Master of Science in Physics from University of California, Los Angeles and a Bachelor of Science in Physics from the California Institute of Technology.

11. I have studied, taught, practiced, and conducted research involving the use of genomic technologies, including DNA sequencing and DNA microarrays, and involving computational analysis of data related to genomic technologies. I have

expertise in genomic technologies, computational analysis, and modeling of a plant, animal, and microbial systems with an emphasis on the study of human diseases, such as cancer.

12. In describing my work, I often use terms to refer to techniques and technologies, as well as specific application areas. During the course of my career, I have worked in genomics, functional genomics, epigenomics, bioinformatics, biostatistics, computational biology, systems biology, network biology, and network medicine. Although the boundaries between these various fields are often not well defined, one possible distinction is that some are directed toward data generation while others are directed toward developing and applying methods for data analysis and interpretation. For example, genomics, functional genomics, and epigenomics are generally associated with the design and conduct of experiments for generation of large-scale data on hundreds, thousands, or tens of thousands of genes, RNA transcripts, or epigenetic sites within the genome (or other alterations such as identifying regulatory interactions or structural changes in the genome). Fields including bioinformatics, biostatistics, computational biology, systems biology, network biology, and network medicine generally involve the development of new methods for data analysis and application of those methods to the biological and biomedical data sets, including those generated using techniques such as genomics, functional genomics, and epigenomics. The beauty of all of these techniques is that

their development and use is, largely although not entirely, independent of their application area. So methods developed on one area are often used in other areas.

13. I have worked in the analysis of genes in hundreds of animal, plant, unicellular, and parasitic species, including human, mouse, rat, cattle, sheep, brown bear, salmon, corn, tomato, rice, yeast, the causative agent of malaria, and countless bacteria and viruses. Despite this diversity of applications, my primary research emphasis has been in the study of human diseases, with the greatest body of my recent work in the analysis of human cancers and pulmonary diseases. However, I have also explored a wide range of other issues in health and disease, including Alzheimer's disease, Parkinson's disease, Huntington's disease, neonatal sepsis, and Down syndrome. My work has involved analysis of data generated from bulk tissue samples, small numbers of cells, single cells, and cell-free DNA.

14. I have also done work relevant to reproductive health, an area that became personally important to me as my wife and I, then in our forties, became enmeshed in the use of assisted reproductive technology. This gave me first-hand exposure to a wide range of techniques, including pre-implantation prenatal screening and the use and limitations of technologies directed toward prenatal detection of genetic abnormalities including aneuploidies. Following the birth of my son in 2006 at Brigham and Women's Hospital in Boston, I also worked with, and mentored, physician scientists at Harvard, published a genomic analysis of

diminished ovarian reserve, and presented the President's plenary lecture at the 63rd American Society for Reproductive Medicine Annual Meeting in 2007 in which I described the use of genomic technology in reproductive, prenatal, and neonatal health.

15. I have received numerous awards and honors including being recognized as a White House Open Science Champion of Change, an Australia Fellow, and a Distinguished International Advisor for the International Conference of BioInformatics and BioEngineering sponsored by the Institute of Electrical and Electronics Engineers. In 2022 I was elected to the National Academy of Medicine, one of three academies that make up the National Academies of Sciences, Engineering, and Medicine (the National Academies) in the United States.

16. I have published more than 300 peer reviewed scientific papers that have been cited more than 89,000 times. I have had continuous grant support for my research from the National Institutes of Health and other organizations for nearly thirty years. In 2018, I received an NCI Outstanding Investigator Career Award, a seven-year research grant recognizing a lifetime record of substantial contributions to the field.

17. A copy of my Curriculum Vitae, attached as Exhibit 1023, contains further details on my education, experience, publications, and other qualifications to render an expert opinion in this matter.

II. Legal Standards

18. I am not a lawyer. However, I have been informed by counsel of general principles concerning patent law and the analysis of whether a patent claim is valid. I have followed these principles in arriving at my conclusions in this declaration. The sections below set forth my understanding of the legal principles that are relevant to my opinions in this declaration.

A. Claim Construction

19. I understand that the claims of a patent define the scope of the inventions that are protected by the patent. I understand that the language of the claims is construed as it would be understood by a hypothetical person having ordinary skill in the art (a “POSA”) at the time of the invention. Accordingly, where I refer below to a POSA—for example, where I discuss how such a person would have understood the specification or the claims, or what a POSA might have known, among other things—I am referring to such a person working at the time of the invention, even if for the sake of brevity I omit that specific timeframe from my statements.

20. I understand words in a claim will be given their ordinary or customary meaning unless it appears that the inventor used them differently.

21. I also understand that, in construing patent claims, courts seek to determine what a person skilled in the art would have understood the claims to mean

primarily in light of the intrinsic evidence of record, which includes the language of the claim itself and other issued claims, the specification, the art cited during prosecution, and the prosecution history.

22. As discussed below in Section VII (“Claim Construction”), where relevant, I have applied the implicit interpretation of the claim language that Patent Owner applied in its complaint for infringement against Petitioner. Ex. 1107.

B. Anticipation

23. I understand that, for an invention claimed in a patent to be found patentable, or to be valid, it must be, among other things, new in light of what came before it. That which came before is generally referred to as “prior art.” I understand anticipation requires that each and every element of a challenged claim is present in or otherwise taught by a single prior art reference.

C. Obviousness

24. Also, I understand that, for an invention claimed in a patent to be found patentable, or to be valid, it must be, among other things, not obvious in light of what came before it, *i.e.*, prior art.

25. I understand that a patent claim is unpatentable or may be found invalid as obvious if, at the time when the invention was made, the subject matter of the claim, considered as a whole, would have been obvious to a person having ordinary

skill in the field of the technology (the “art”) to which the claimed subject matter belongs.

26. I understand that the obviousness standard is defined in the patent statute as follows:

A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 [in a single prior art reference], if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

35 U.S.C. § 103(a).

27. I understand that a claim in a patent is obvious when the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which the subject matter pertains.

28. I understand that obviousness may be shown by considering more than one item of prior art, and that the following factors should be considered in analyzing obviousness: (1) the scope and content of the prior art; (2) the differences between

the prior art and the claims; (3) the level of ordinary skill in the pertinent art; and, if in evidence, any secondary considerations of non-obviousness.

29. I understand that a person of ordinary skill in the art is assumed to have knowledge of all relevant prior art. I understand that one skilled in the art can combine the teachings of various prior art references based on the teachings of the individual prior art references, the general knowledge present in the art, or common sense. I understand that a teaching, motivation or suggestion appearing in the prior art may provide a basis for combining two (or more) prior art references or modifying a given prior art reference. A motivation to combine references may also be implicit in the prior art, and there is no requirement that there be an actual or explicit teaching to combine two references. Thus, one may take into account the inferences and creative steps that a person of ordinary skill in the art would employ to combine the known steps in the prior art in the manner claimed by the patent at issue.

30. I understand that the combination of familiar steps according to known methods is likely to be obvious when it does no more than yield predictable results. I understand that when a work is available in one field of endeavor, design incentives and other market forces can prompt variations of that work, either in the same field or a different one. Likewise, if a technique has been used to improve one device, and a person of ordinary skill in the art would recognize that it would improve similar

devices in the same way, using that technique to improve the other device in a predictable manner would normally be obvious.

31. I understand that a person of ordinary skill is also a person of ordinary creativity in the field. A person of ordinary skill attempting to solve a problem will not be restricted only to those steps of prior art designed to solve the same problem. Rather, in exercising common sense and ordinary creativity, a person of ordinary skill can draw upon prior art directed to any need or problem known in the field of endeavor at the time of the invention. In other words, the prior art does not need to be directed towards solving the same problem that is addressed in the patent. Further, the individual prior art references themselves need not all be directed towards solving the same problem.

32. I further understand that in many fields, it may be that there is little discussion of obvious techniques or combination, and it often may be the case that market demand, rather than scientific literature or knowledge, will drive design trends. When there is such a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within their technical grasp. If this leads to the anticipated success, it is likely the result not of innovation but of ordinary skill and common sense. In that instance, the fact that a combination was obvious to try might show that it was obvious. I understand that the fact that a

particular combination of prior art steps was “obvious to try” may indicate that the combination was obvious even if no one had attempted the combination at the time of filing. Likewise, duplicating a known feature from the prior art multiple times in order to take advantage of proven technology typically does not require a leap of inventiveness.

III. The ‘968 Patent

33. The ‘968 patent (Ex. 1001) is titled “Methods for Using Mosaicism in Nucleic Acids Sampled Distal to Their Origin.” Ex. 1001, Code (54). The patent claims priority to “[p]rovisional application No. 62/072,936, filed on October 30, 2014.” *Id.*, Code (60). I have considered October 30, 2014 to be the relevant date for assessing prior art invalidity in my analysis, although I offer no opinion on whether the ‘936 provisional application in fact provides sufficient support for the issued claims of the ‘968 patent.¹

IV. Person of Ordinary Skill in the Art

34. I have been advised that a “person of ordinary skill in the art,” or “POSA,” is a hypothetical person and that I should consider the following factors in constructing a POSA: (1) the educational and experience level of active workers in the field; (2) the sophistication of the technology; (3) the rapidity with which

¹ I am not aware of any evidence suggesting that Patent Owner is entitled to an earlier priority date.

innovations are made; (4) the types of problems encountered in the art; (5) the prior art solutions to those problems. I understand that not all such factors may be present in every case, and one or more of them may predominate.

35. The relevant technology field for the '968 patent is nucleic acid sample preparation, sequencing, and analysis to detect genetic variation for clinical applications. By virtue of my education, experience, and training, I am familiar with the level of skill in the art of the '968 patent before October 30, 2014.

36. After considering the factors above, I believe a person of ordinary skill in the relevant field before October 30, 2014 would have had an advanced degree (such as an M.D. or the equivalent of a Ph.D. in a life or physical sciences, engineering, or mathematics discipline), and through further education or postdoctoral experience, an understanding of various sequencing and next generation sequencing methods, sample preparation methods, targeted enrichment methods, bioinformatics methods for interpreting sequencing data, and methods for identifying genetic variants in a sample. A POSA also would have had an understanding of clinical applications of sequencing technology to conditions associated with nucleic acid variation, such as cancer, non-invasive prenatal testing, and transplant rejection. Additionally, a POSA would have had an understanding of cancer biology as it relates to mutations associated with tumor diagnosis,

progression, and treatment, as well as awareness of the various types of mutations that can occur in and influence the course of a patient's cancer.

37. I understand that a POSA is presumed to be aware of all relevant art.

V. Overview of a Person of Ordinary Skill in the Art's Knowledge

38. In this Section, I provide an overview of the knowledge that a POSA would have possessed as of October 30, 2014.

A. Overview of Molecular Biology and Genetics

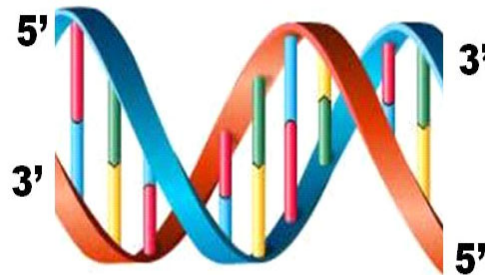
39. I begin with an overview of the principles of biology that are relevant to understanding the state of the art, especially those pertaining to nucleic acid sequencing and cancer biology.

1. The Central Dogma of Molecular Biology

40. Genes contain the set of instructions, or blueprints, for the development and maintenance of the cells of all organisms. Genes are made of deoxyribonucleic acid ("DNA") and are located on chromosomes in the nucleus of eukaryotic cells. DNA is composed of subunits called nucleotides, the sequences of which generate the code directing the eventual production of proteins.

41. DNA exists as a double-stranded molecule: it consists of two long strands in a double-helical structure. One end of each strand is called the 5-prime ("5'") end, and the other is called the 3-prime ("3'") end. The strands are bound

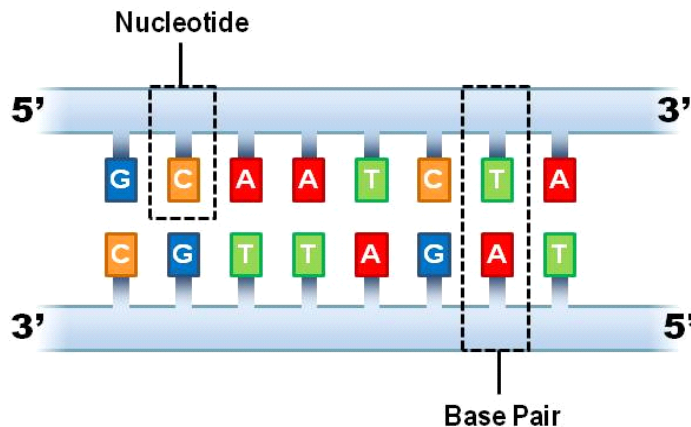
together such that they run in opposite directions: one strand, oriented in the 5' to 3' direction, is attached to another strand, oriented in the 3' to 5' direction.



42. The smallest functional unit of DNA is the gene, which is a sequence of nucleotides that encode for ribonucleic acid (“RNA”), which in turn codes for a protein. Proteins are incredibly versatile large molecules that serve crucial functions in essentially all biological processes. As examples: some proteins, called enzymes, serve as catalysts for the creation of other biological molecules; some transport and store other molecules, such as oxygen, throughout the body; some provide immune protection; some provide physical support and generate movement; and others transmit nerve impulses.

43. Attached at regular intervals along the backbone of each strand of DNA are “bases.” Each base, with its corresponding piece of DNA backbone, is referred to as a “nucleotide.” A strand of multiple nucleotides is called an “oligonucleotide” or “oligo.”

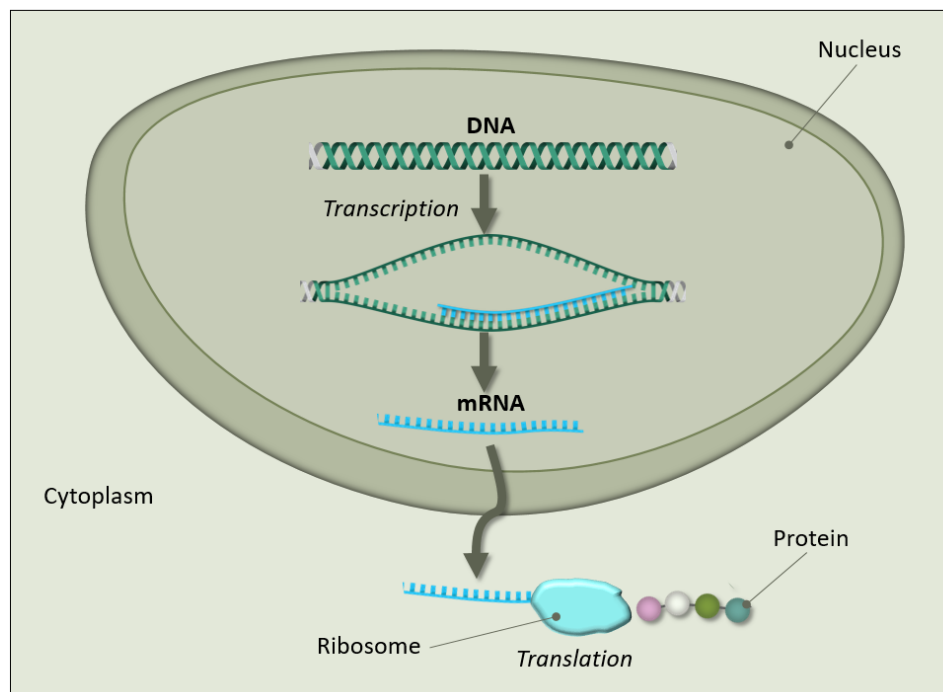
44. There are four natural types of bases that occur in DNA: adenine (“A”), cytosine (“C”), guanine (“G”) and thymine (“T”). Each type of base on one DNA strand binds to only one type of base on the other strand, known as its “complementary” base. When complementary bases bind together, or “hybridize,” they form what are known as “base pairs.” A hybridizes to T, and G hybridizes to C. It is the pairing, or hybridization, of the complementary bases that holds the two strands of DNA together, as depicted below.



45. The A-T and G-C interactions between strands of oligonucleotides are themselves specific. However, it is possible for two strands to bind together even if not all of the bases in them are perfectly complementary. That is, as long as a sufficient number of A-T and G-C pairs can form between two oligonucleotides, they will “anneal” or “hybridize,” even if there are some mis-matches within the two strands.

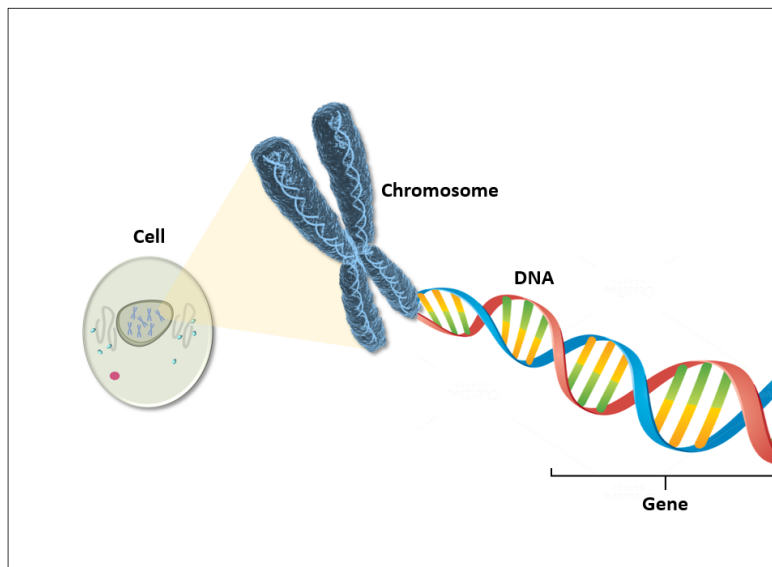
46. The information encoded by the DNA in a gene is turned into a protein via two step-wise processes, called transcription and translation. In transcription, the DNA is used as a template by specialized machinery in the cell to generate another type of molecule, called messenger ribonucleic acid (mRNA, that contains a series of nucleotides “complementary” to those in the DNA-encoded gene). In translation, the mRNA is read by an organelle called the ribosome three nucleotides at a time. These three-nucleotide mRNA strings, called codons, instruct the addition of protein building blocks called amino acids one-by-one by the ribosome’s translational machinery.

47. The flow of information from DNA to RNA to protein is referred to as “the central dogma of molecular biology.”



2. Genetics

48. The human genome consists of approximately 3.2 billion nucleotides of DNA, and (most) human cells contain two copies of the genome, one inherited from each biological parent. The DNA in the genome is divided and packaged among the chromosomes. Humans have 23 chromosome pairs: 1 pair of sex chromosomes (XX—female; XY—male), and 22 autosome pairs (the chromosomes that are not sex chromosomes). Encoded within each copy of the genome are 25,000-30,000 genes, each located at a specific location on one of the chromosomes.



49. Although only about 1% of the overall DNA in humans is made up of protein-encoding genes, variation in the sequences of these genes is largely responsible for the differences between us. While a person's genes are important for proper biological function, no two people (apart from identical twins) have identical DNA sequences through the entirety of their genomes. Certain alterations,

sometimes called “mutations,” “variations,” “variants,” or “polymorphisms” can occur in all genes and portions of the human genome, contributing both to normal patterns of human variation as well as individual risk for developing diseases. Polymorphisms can include many types of genomic alteration, including insertions, deletions, point mutations, translocations, rearrangements, and copy number variations. The entire collection of an individual’s genetic variants is referred to as their “genotype.”

50. A specific DNA sequence is said to be polymorphic if it varies between individuals, and the individual sequence variants are referred to as alleles. Each allele can have varying influences on the protein encoded by the gene. Because of redundancies in the genetic code, two slightly different alleles may produce exactly the same protein. Other alleles may cause a change in the protein encoded by a particular gene, altering the protein’s function. Still others may include a mutation that instructs the cell’s translational machinery to prematurely stop, resulting in the production of a truncated protein with complete loss-of-function. Other allelic variations may have resultant effects that fall somewhere in-between, with different amino acid sequences having varying levels of effect on protein function.

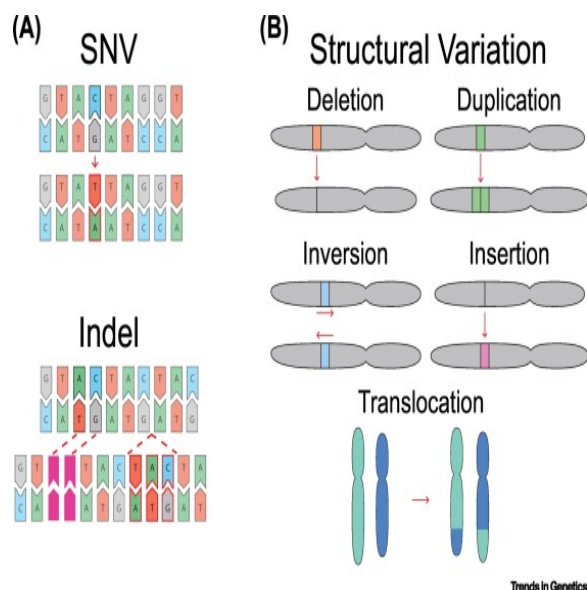
51. In the autosomes (the non-sex chromosomes), there are two alleles for each gene, one coming from each parent. When each of the alleles are the same, the

individual is said to be homozygous at the allele, and when the alleles are different, the individual is said to be heterozygous at the allele.

B. Genetic Mutations

52. Genetic mutations can be inherited (*i.e.*, passed on from parents to their offspring) or acquired (*i.e.*, developed for the first time in a new cell). Inherited mutations are called “germline” mutations. Acquired, or non-inherited, mutations are called “somatic” mutations. These can result from chance cellular events, such as errors acquired during cell division, as well as from environmental and lifestyle factors that induce damage to the genome, such as exposure to carcinogens. Somatic mutations are not present in all of the cells in one’s body; they are only present in the progeny of the cell in which the mutation first occurred. Those skilled in the art appreciated that, unlike germline mutations that are inherited and present in every cell in the body, somatic tumor mutations arise during a person’s lifetime and occur only in the cells that have multiplied from a cell in which the somatic mutation has originated. Commonly, such mutations may first arise in cancerous (or tumor) cells, and thus propagate further as the cancer rapidly divides. Often, mutations that appear early in tumor development were likely to present in the tumor’s cells throughout the course of its development. Thus, somatic mutations unique to a particular tumor can be a useful tool for detecting, tracking, and monitoring a patient’s cancer.

53. Different types of somatic mutations were known in the prior art well before October 2014. *See* Ex. 1024 (<https://www.garvan.org.au/research/kinghorn-centre-for-clinical-genomics/learn-about-genomics/for-gp/genetics-refresher-1/types-of-variants>). These include single nucleotide polymorphisms (“SNPs”), copy number variants (“CNVs”), and insertions or deletions (“indels”). An SNP (sometimes called a single nucleotide variant (“SNV”)) is a change in one DNA base pair that may result in the substitution of one amino acid for another in the protein that is encoded by the gene. The effect that an SNP has on the protein’s function can vary widely, from having no effect at all to having a very significant effect. CNVs can result in producing too much or too little of a protein, which can have significant impacts on cell function. And indels can include the insertion of extra base pairs into a gene and/or the deletion of normal base pairs from the gene, which can also significantly affect the functioning of the proteins that the affected gene encode.



54. A POSA would have also been familiar with a type of polymorphism known as a translocation, which can move DNA from one chromosome to another chromosome (interchromosomal translocation) or to a different position on the same chromosome (intrachromosomal translocation). This type of polymorphism was known to be particularly relevant to a class of blood cancers called leukemias.

55. By 2014, there were also publicly accessible databases that a POSA would have been aware of, that could be searched for a large number of catalogued cancer-associated mutations. These included databases maintained by the International Cancer Genome Consortium (“ICGC”) (*e.g.*, the IARC/WHO TP53 database), The Cancer Genome Atlas (“TCGA”), International Agency for Research on Cancer (“IARC”), and the Catalogue of Somatic Mutations in Cancer (“COSMIC”).

C. Nucleic Acid Sequencing Technologies

1. Sequencing Techniques

56. Nucleic acid sequencing is the science of determining the order of the bases in a nucleic acid molecule on a base-by-base level. DNA sequencing requires a minimum amount of DNA. Thus, before it can be sequenced, a sample with too little DNA generally must be amplified.

57. One of the earliest known sequencing methods is called Sanger sequencing. Sanger sequencing is carried out by replicating a template strand of

DNA, and incorporating labeled, chain-terminating bases at different locations along the different growing strands. The DNA fragments are then separated by size using electrophoresis and the various lengths of the fragments are used with the terminating bases to reconstruct the whole sequence of the template.

58. In Sanger sequencing, single-stranded DNA “templates” (*i.e.*, the DNA to be sequenced) are placed in solution containing a polymerase, oligonucleotide primers (to start the reaction), and deoxynucleotides (for base pairing and primer extension). Sanger sequencing relies on the use of, in addition to regular deoxynucleotides, a small admixture of synthetic di-deoxynucleotides, which are labeled, for example, with fluorescent tags of different colors indicative of the particular base—A, G, T, or C—and serve a “stop” function as explained below.

59. During strand extension in Sanger sequencing, when a normal deoxynucleotide is added to the growing strand, the next position remains “open” for another deoxynucleotide or di-deoxynucleotide to pair up to the template and elongate the copy strand. Once a di-deoxynucleotide is incorporated into the growing strand, it ends there, and no new nucleotides can be added on. Because the terminal di-deoxynucleotides are labeled, one can detect the last base that was added onto the growing strand. But if one has hundreds or thousands of copies of each strand, then during the sequencing process each copy of the same original DNA strand will see

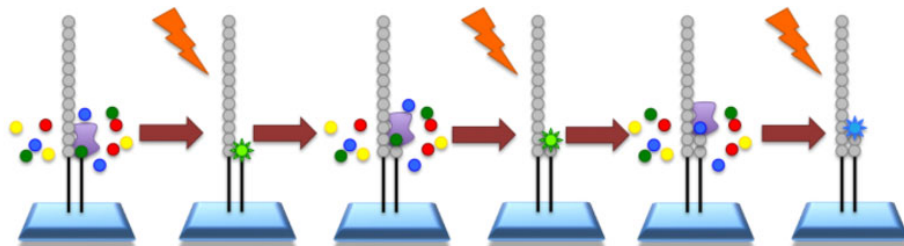
a di-deoxynucleotide incorporated randomly, leading to fluorescently labeled strands of different lengths.

60. All of the fluorescently labeled copies of different lengths are then separated based on their size by a process called gel electrophoresis. This creates a fluorescent “read” of the sequence from the shortest strands (those terminating immediately after the primer sequence) to the longest strands (those terminating after being extended all the way to the end of the template). A computer processes the fluorescent data and produces a “trace file” that displays a representation of the fluorescent signal and the corresponding DNA sequence.

61. By 2005, a new approach to sequencing often referred to as “next generation sequencing” (“NGS”), sometimes also referred to as “massively parallel sequencing,” or “sequencing by synthesis” had been developed to allow “multiplex” sequencing of hundreds of thousands or millions of different DNA molecules simultaneously. By approximately 2008, techniques for NGS were well-known and commercially available. *See, e.g.,* Ex. 1013 (Shendure 2008).

62. NGS is a high-throughput method for determining the base-by-base sequences of nucleic acid molecules, such as DNA. NGS is exponentially faster, cheaper, and more efficient than the Sanger sequencing technique that preceded it. For example, using NGS, the entire human genome can now be sequenced on a single instrument in a single day. In NGS, the DNA molecules being sequenced are

first modified to attach what are often referred to as “sequencing adapters” to each end of the molecules. These modified DNA fragments are then rendered single-stranded and allowed to attach at low density to a “flow cell” in which the individual fragments can be subjected to “bridge amplification,” a PCR-like technique that creates molecular “clusters” that facilitate sequencing and detection. The DNA molecules are again rendered single-stranded so that their nucleotides can be used as a template for DNA polymerase to make a copy. The polymerase builds a complementary strand in which single labeled nucleotides are added, illuminated with a laser and detected because they glow in a color corresponding to the base (A, G, T, or C) at that position, and then “deprotected” by releasing the fluorescent dye so the next labeled nucleotide can be added. As this process progresses and the complementary strand grows in length, the pattern of colors reveals the identity of the string of nucleotides, which is recorded by a computer that uses the series of detected bases to reconstruct the sequence. Hence the sequence is determined through the synthesis of a new (complementary) strand. This is shown in the figures below.



63. A computer program can be used to align the sequence read back to a reference genome, or another sequence reference, which locates the sequence in the genome and thereby determines the identity (*i.e.* in which gene or on which chromosome) of the sequenced nucleic acid.

2. Sequencing Depth and Read Counts

64. NGS has been applied widely to sequencing of genomic sequences. It was known well before October 2014 that different applications would involve different levels of sequencing, or sequencing depths. Generally, to obtain highly accurate sequencing information, it is necessary to sequence accessible parts of the target genomic sequences multiple times. Reasons for this include (i) multiple observations per base are required to have confidence that the base “call” is correct, so that errors that might occur in any one sequencing read can be rejected; (ii) so that mutation calls can be made with confidence based on multiple independent sequencing reads exhibiting the same polymorphism; and (iii) since DNA is randomly fragmented before sequencing and the amplification and library construction process may introduce some biases, sequencing reads are not evenly distributed throughout the genome, meaning that some bases will be read more than other bases. Ex. 1028 (Illumina Technical Note, “Estimating Sequencing Coverage”).

65. “Coverage,” “sequencing depth,” or “depth of sequencing” are terms that refer to the number of times genomic sequence has been sequenced. It can be calculated by the Lander/Waterman equation using the formula:

$$C=L*(N/G)$$

Where C stands for coverage, G is the haploid genome length (~3 billion basepairs for the human genome), L is the read length, and N is the number of reads. After the initial “completion” of the human genome project (1990-2003) in April 2003, a draft human genome sequence was produced, which served as a framework upon which future human genome sequences could be overlaid and interpreted. The advent of NGS accelerated this effort dramatically in the years that followed. Initial efforts to produce the whole genome sequences of specific individuals including those of Dr. James Watson (of double-helix fame) and Dr. J. Craig Venter (of human genome project fame) in 2007 reported use of NGS to target coverage depths of 7.4-7.5x. However, a subsequent study in 2008 revealed that higher coverage depth (~30x) allowed for improvements in the fidelity of the sequences, as relevant for polymorphism discovery. Ex. 1077. By 2010, whole-genome sequencing to 30x coverage was routine on the commercially available Illumina HiSeq 2000 (I purchased one of these devices for us in my own lab in 2009 with funding from an NIH Shared Instrument Grant). Ex. 1029 (Illumina Specification Sheet, 2010) (“With unmatched cost-effectiveness, HiSeq 2000 is the first commercially available

sequencer to enable researchers to obtain ~30x coverage of two human genomes in a single run for under \$10,000 USD per sample.”; “The ultra-high output and speed of HiSeq 2000 now make it possible to sequence two human genomes at ~30x coverage simultaneously”). One run of the HiSeq 2000 allowed for a billion single-end reads or two billion paired-end reads (sequencing reads from both ends of a DNA fragment). *Id.* at 2. As each read was 100 bases in length, for 2 billion reads, the formula above yields 66.67x coverage for a single genome or 33.33x coverage of two genomes.

66. Whole genome sequencing experiments that sought to ensure more complete coverage of the genome could do so by sequencing to greater depth. For example, in 2012, Lam reported performing whole genome sequencing to an average coverage of ~76x using the HiSeq2000 and a different sequencing platform from Complete Genomics. Ex. 1027 at 78. Sequencing to even greater depth was also routine, as it only required additional runs of the sequencer: “it is very easy to increase the coverage or sequence depth, if you later decide you need more data. Provided you still have your original sample, you can just sequence more, and combine the sequencing output from different flow cells.” Ex. 1028 at 2.

67. In February 2014, Sims reviewed many published sequencing studies. Ex. 1040. In particular, Sims reviewed all the sequences from the Illumina platform deposited from 2012 to June 2013 by 771 different studies into the European

Nucleotide Archive (<https://www.ebi.ac.uk/ena>). Sims found that the number of bases sequenced per sample for whole genome sequencing and whole exome sequencing varied widely. For whole genome sequencing, the range was about 0.1 GB to over 200 GB. *Id.* at 123, Figure 1. For whole exome sequencing, the range was about 0.2 GB to 20 GB. *Id.* Indeed, scientists chose the number of bases sequenced based on the goal of the particular sequencing project.

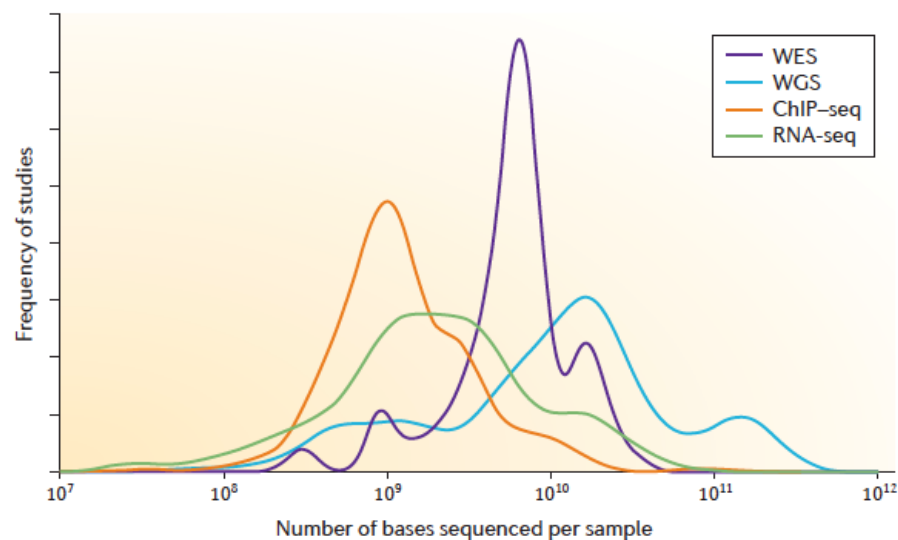


Figure 1 | Sequencing depths for different applications. The frequency of studies that use read counts of all runs (which are typically flow-cell lanes) and that were deposited from 2012 to June 2013 for the Illumina platform in the European Nucleotide Archive (ENA) is shown. The plot provides an overview of sequencing depths that are usually chosen for the four most common experimental strategies. Densities have been smoothed and normalized to provide an area under the curve that is equal to one. The depth and therefore the cost of an experiment increase in the order of chromatin immunoprecipitation followed by sequencing (ChIP-seq), RNA sequencing (RNA-seq), whole-exome sequencing (WES) to whole-genome sequencing (WGS). Although ChIP-seq, WES and WGS have typical applications and thus standardized read depths, the sequencing depth of RNA-seq data sets varies over several orders of magnitude. Multimodal distributions of WES and WGS reflect different target coverage. To generate this figure, runs were summed by experiment and, for each study, one experiment was chosen at random to avoid counting large studies more than once. Note that the ENA archive only contains published data sets and excludes medically relevant data sets. The plot was created from 771 studies.

3. Amplification and Enrichment

68. Both Sanger sequencing and NGS require minimum quantities and/or relative abundances of target DNA. Thus, well before 2014 several techniques had been developed to generate or isolate copies of DNA molecules of interest to facilitate sequencing.

69. DNA enrichment is the process of increasing the concentration of some DNA relative to other DNA in a mixture. This can be done using biochemical methods that involve physically pulling certain pieces of DNA out of a mixture for further analysis. By the early 2000s, enrichment of desired targets on solid substrates like DNA microarrays was in use. Biochemical approaches to enrichment can involve using oligonucleotides with sequences complementary to the target DNA, which allows them to hybridize to other oligonucleotides that one seeks to analyze in a mixture. These oligonucleotides are referred to sometimes interchangeably as primers, probes, or baits.

70. Enrichment generally involves increasing the amounts of specific targets to generate sufficient quantities of a nucleic acid sequence to detect and analyze. *See, e.g.,* Ex. 1007 (Mamanova 2010) (“We have not yet reached a point at which routine sequencing of large numbers of whole eukaryotic genomes is feasible, and so it is often necessary to select genomic regions of interest and to enrich these regions before sequencing.”). I focus here on PCR and solution-based hybrid capture

techniques as they are the most pertinent to the prior art references I have used in analyzing the grounds of unpatentability. However, a POSA would have been familiar with several other related techniques, such as the use of molecular inversion probes (“MIPs”), array-based systems, and solid-phase hybrid capture, among others. *See id.* (describing PCR and solution-based hybrid capture as among “the most widely used approaches to target enrichment”); *see also id.* at Fig. 2 (reproduced below) (showing overlapping applicability of target-enrichment strategies).

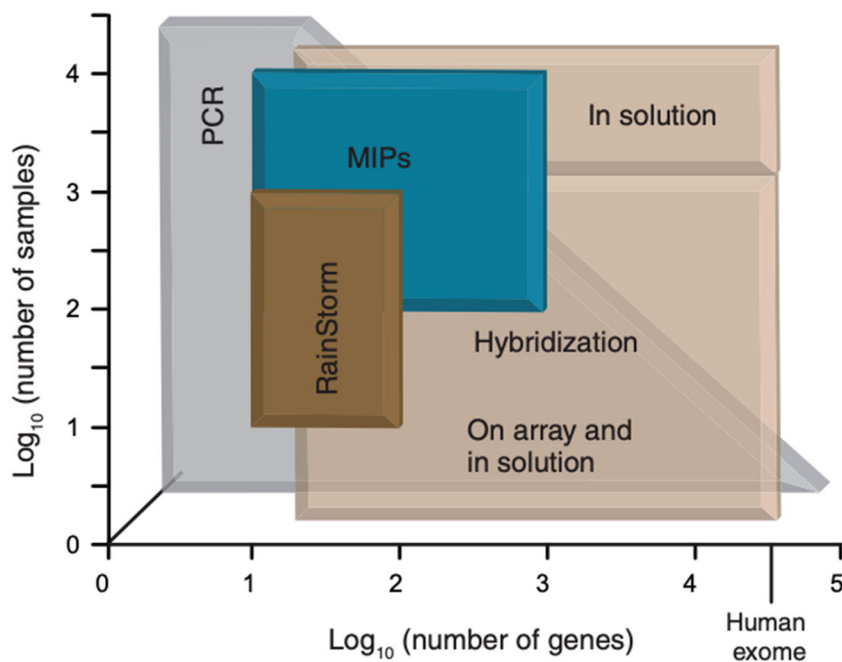


Figure 2 | Suitability of different target-enrichment strategies to different combinations of target size and sample number. Suitability was estimated from the perspective of the feasibility with which each method could be applied to the various combinations of target size and sample number, rather than the cost.

71. While these strategies have considerable overlap in their possible applications, a fair summary of the advantages of each technique is set forth in Mertes (Ex. 1006):

Overall, one can draw the following conclusions from a comparison of the currently used enrichment techniques : (i) that hybrid capture has its main advantages for medium to large target regions (10-50 Mb) in contrast to the other two approaches which typically only target small regions within the kilo base pairs and low mega base pairs range. The ability to enrich for mega base pair-sized targets is particularly advantageous in research studies where typically whole exomes or many genes are involved. Especially for clinical applications, this may be relevant for oncological applications where one would expect to sequence 100-1000's of genes. (ii) The advantage of PCR and circularization-based methods [i.e., MIPs] is that they achieve very high enrichment factors and few off-target reads, but only for small target regions. This is more suited to clinical genetics where typically only few critical loci need to be assessed.

Ex. 1006 (Mertes 2011) at 380. Thus, in assays where a POSA sought to analyze many sequences of interest, solution-based hybrid capture would have been an ideal enrichment strategy. On the other hand, in assays where a POSA intended to analyze relatively few sequences of interest, then a PCR approach would have been particularly well-suited. For intermediate numbers of potential targets, either technique might serve equally well. However, these parameters would not have been

considered as strict limitations, and the techniques can and were used on a relatively interchangeable basis in many experimental protocols, and could also be used together as complements to one another. *See* Ex. 1006 (Mertes 2011) (Table 1) (describing pros, cons, and overlapping applications of “[c]urrently employed targeted enrichment techniques”).

Table 1: Currently employed targeted enrichment techniques

Enrichment technique	Vendor	Features	Pros	Cons	Number of loci (target size)	Library prep for NGS
Hybrid capture						
Solid support	Agilent SureSelect, Roche NimbleGen SeqCap EZ	Medium to large target regions, custom and preconfigured target regions (i.e. whole exome), Multiplexing possible, ready to use kits	Ease of production, large target sets	Large amount of input DNA, high-tech equipment (3–10 µg)	10^4 – 10^6 (1–50 Mb)	Before enrichment
In solution	Agilent SureSelect, FlexGen FlexSelect, Mycroarray Myselect, Roche NimbleGen SeqCap EZ		Ease of use, small amount of input DNA (<1–3 µg)			
Circularization						
Molecular inversion probes						
Selector probes	HaloGenomics	Custom kits and clinically relevant panel kits	No dedicated instruments, high specificity, input DNA (<1 µg)	exome kit not available yet	10^2 – 10^4 (0.1–5 Mb) 10 – 200 (0.1–1.5 Mb)	During enrichment (incorporated into hybridization probes)
PCR						
Long range	Invitrogen SequalPrep, Qiagen SeqTarget system	Smaller target regions, coverage by tiling	Relatively easy to set up and automatable, even coverage	PCR conditions largely influence effectiveness, >10 µg DNA for Large sets	10^2 – 10^4 (0.1–5 Mb)	After enrichment
Multiplex	Multiplicom, Fluidigm	Smaller target regions, coverage by tiling, multiplex PCR of 150–200 amplicons (150–450 bp)	Easy to perform, reasonably economical in terms of, even coverage			
Micro droplet	RainDance	Smaller target regions, coverage by tiling, micro droplet PCR of up to 20 000 amplicons (150–1500 bp)	Even coverage, low input DNA	Relatively expensive, specialist equipment	10^3 – 10^4 (up to 10 Mb)	

All major targeted enrichment techniques show relative pros and cons.

a) PCR

72. In PCR, DNA molecules are replicated in a manner that takes advantage of the base pairing of complementary DNA strands.

73. In PCR, the DNA “template” strands (*i.e.*, the strands being replicated) must first be separated. This is because DNA normally exists in a double-stranded form, but it must be single-stranded to be replicated. Thus, the strands of double-

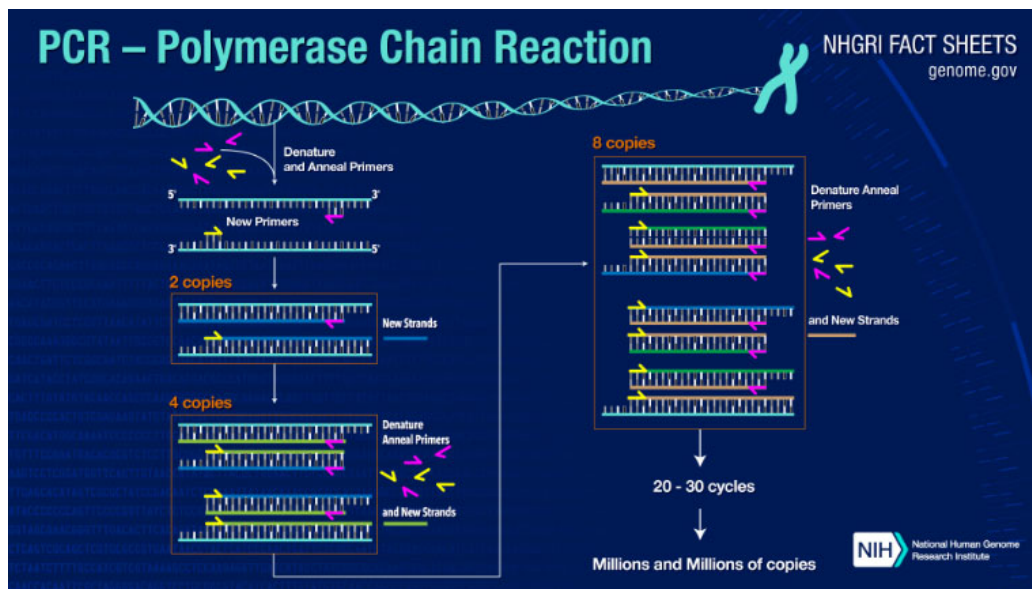
stranded DNA are first completely separated (or “denatured” using heat, or other adjustments to the conditions of the reaction), leaving two complementary single-stranded DNA template strands.

74. PCR “primers” or “probes,” which are short oligonucleotides that are complementary to a small portion of the DNA templates, can then be added to the denatured DNA at high concentrations. These primers/probes then hybridize (*i.e.*, bind through base pairing interactions) to the DNA template strand at a particular target region of interest (specified by the order of bases) to capture that specific sequence for replication. The small portions of the template to which the primers/probes hybridize are referred to as the “priming sites.”

75. After the PCR primer hybridizes to the target sequence, an enzyme called a “polymerase” then synthesizes a new nucleic acid molecule complementary to the single-stranded DNA template starting at the 3’ end of the primer. The polymerase then adds nucleotides (one nucleotide at a time) complementary to the rest of the template until a complete copy of the template is made.

76. After a complete copy is made, the strands are then denatured (again, typically through heating), and the process starts over, each time allowing for the doubling of the copies of template strands present in a sample following hybridization of the PCR primers to the priming sites. Each round of denaturation, primer hybridization, and extension is referred to as a “cycle.”

77. PCR results in the exponential amplification of target DNA sequences. In the first cycle, two new strands of DNA are created in addition to the template strand of DNA. In the second cycle, copies are made of each of the four strands, for a total of eight strands. As PCR continues, the number of copies increase at an exponential rate as depicted in the schematic below.

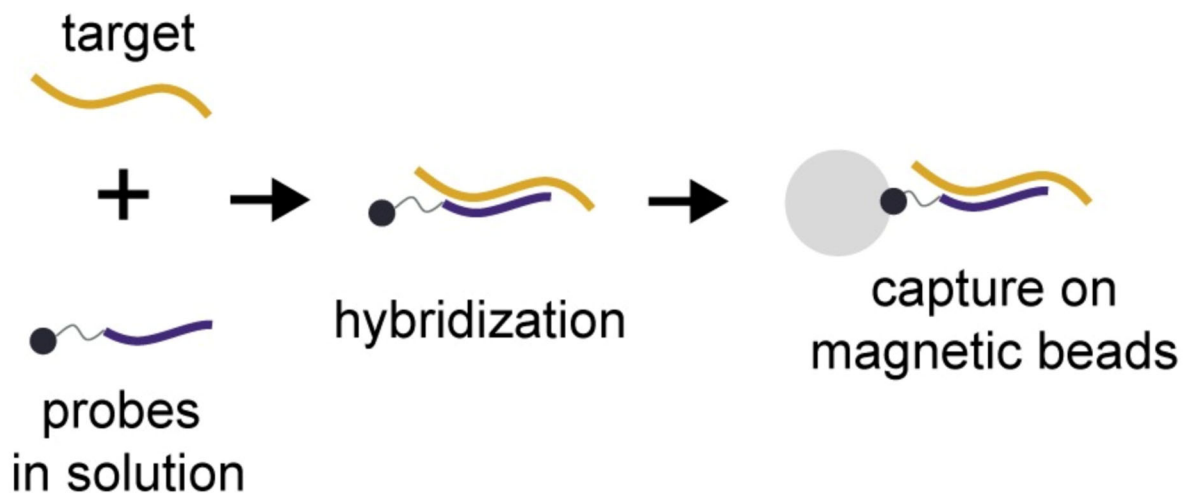


b) Solution-Based Hybrid Capture

78. A POSA would have also been familiar with the use of solution-based hybrid capture techniques for target enrichment. In short, this technique works by mixing in solution the target DNA with an abundance of short nucleic acid “baits,” which hybridize to the target DNA. The baits also include an additional component (typically biotin) that serves as a ligand for a binding partner (typically streptavidin)²

² The streptavidin-biotin complex is well-known to be one of the strongest binding interactions occurring in nature. The binding strength is significantly greater

that is affixed to a magnetic bead, to allow bead-probe conjugation following the hybridization step. When such beads are mixed into the solution, the target sequence is captured through bead-probe binding, and the target DNA can be isolated by washing the uncaptured DNA and probes.



The key aspects of this process were described in detail in a seminal 2009 article by Andreas Gnirke and Chad Nusbaum at the Broad Institute. Ex. 1005 (Gnirke 2009).

79. When hybrid capture baits are contacted with the nucleic acid sample (the “pool”), the baits hybridize to the sequences of interest. This hybridization results in a “catch” of nucleic acids bearing the targeted sequences, which are “caught” through the hybridization of strand-to-strand base-pairing. The catch is then pulled down or fished out using beads that bind the biotin that was incorporated

than that between two complementary DNA strands, and is far beyond the normal protein-ligand binding strength.

into the baits, and the catch can then be amplified using PCR before sequencing, for example using NGS workflows. Ex. 1005 (Gnirke) at 183; Fig. 1 (reproduced below). *See also* Ex. 1006 (Mertes 2011) at 374 (“By selective recover[y] and subsequent sequencing of genomic loci of interest, costs and efforts can be reduced significantly compared with whole-genome sequencing.”); 375 (“Targeted enrichment can be useful in a number of situations where particular portions of a whole genome need to be analyzed.”).

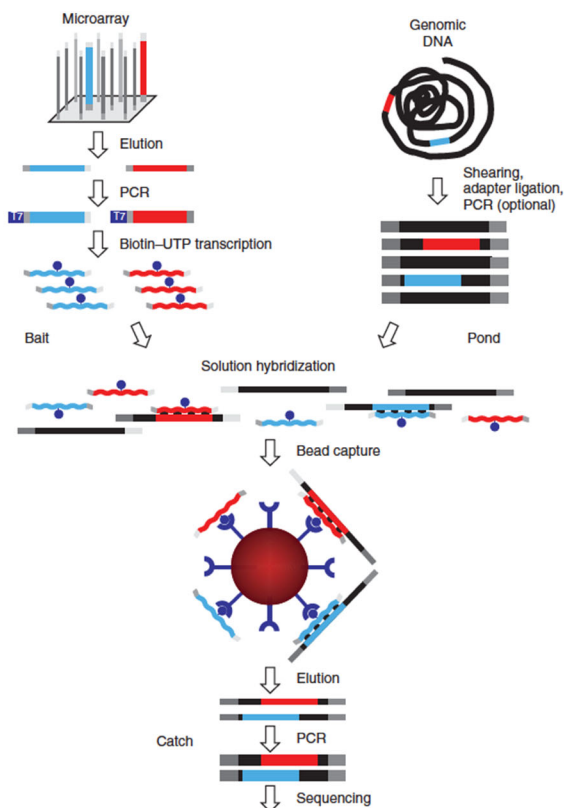


Figure 1 Overview of hybrid selection method. Illustrated are steps involved in the preparation of a complex pool of biotinylated RNA capture probes (bait; top left), whole-genome fragment input library (pond; top right) and hybrid-selected enriched output library (catch; bottom). Two sequencing targets and their respective baits are shown in red and blue. Universal adaptor sequences are gray. The excess of single-stranded nonself-complementary RNA (wavy lines) drives the hybridization.

80. Well before the priority date, “an increasing number of protocols and vendors [had] begun offering out of the box solutions for hybrid capture, meaning, the researcher need not do development work but merely choose between a preset targeted enrichment regions (*e.g.* whole exome) or specify their own custom enrichment region.” Ex. 1006 (Mertes) at 379.

D. Personalized Cancer Testing and Monitoring

81. Once NGS techniques became available, those skilled in the art immediately sought to leverage them to develop methods to discover, monitor, and track mutations in cancer patients. For example, a 2008 *Nature* paper by Timothy Ley and others applied whole-genome sequencing and personalized monitoring techniques to improve treatment of acute myeloid leukemia, the treatment of which “ha[d] changed little in the past two decades, because most of the genetic events that initiate the disease remain undiscovered.” Ex. 1012 (Ley 2008) at 66. Once “[w]hole-genome sequencing [became] possible at reasonable cost and timeframe,” Ley “use[d] this approach for the unbiased discovery of tumour-specific somatic mutations that alter the protein-coding genes.” *Id.* Through sequencing tumor samples that were matched to normal counterparts from skin samples taken from the same patient, Ley “discovered ten genes with acquired mutations; two were previously described mutations that are thought to contribute to tumour progression, and eight were new mutations present in virtually all tumour cells at presentation

and relapse” *Id.* The authors reported that “[their] study establishes whole-genome sequencing as an unbiased method for discovering cancer-initiating mutations in previously unidentified genes that may respond to targeted therapies.”

Id.

82. Ley described the specifics of its protocol as follows:

Our group has pioneered the use of whole-genome re-sequencing and variant discovery approaches using the Illumina/Solexa technology with the genome of the nematode worm *Caenorhabditis elegans* as a proof-of-principle. This approach has distinct advantages in reduced cost, a markedly increased data production rate, and a low input requirement of DNA for library construction. In the present study, we used a similar approach to sequence the tumour genome of a single AML patient and the matched normal genome (derived from a skin biopsy) of the same patient. After alignment to the human reference genome, sequence variants were discovered in the tumour genome and compared to the patient’s normal sequence, to the dbSNP database, and to variants recently reported for two other human genomes; revealing new single nucleotide and small insertion/deletion (indel) variants genome-wide. Somatic mutations were detected in genes not previously implicated in AML pathogenesis, demonstrating the need for unbiased whole-genome approaches to discover all mutations associated with cancer pathogenesis.

Ex. 1012 (Ley) at 67. The whole-genome sequencing was performed using “the Illumina Genome Analyser to achieve the targeted level of 90% diploid coverage as

determined by the coverage of the high-quality SNP set. Maq [a genome assembly software] was used to perform alignment, determine consensus, and identify SNVs within the 98 billion bases generated from the tumor genome.” *Id.* They also “sequenced the genome of normal skin cells from the same patient to enable the identification of inherited sequence variants in the tumor genome.” *Id.*

83. Using the identified mutations, in order “[t]o better define the percentage of tumour cells that contained each of the discovered somatic mutations,” Ley designed PCR primer probes specific to each selected mutation and then “amplified each mutation-containing locus from non-amplified genomic DNA derived from the *de novo* and relapse tumour samples, and from the skin biopsy obtained at presentation.” *Id.* at 69. By looking at the *de novo* (initial) and relapse tumor samples, Ley described analyzing the nucleic acid levels present in the tumor at multiple time points throughout the course of the disease. The resulting enriched populations of these sequences then “were sequenced using the Roche/454 FLX³] platform, and the frequency of reads containing the reference and variant alleles were defined.” *Id.* The allele frequency of the variant alleles is shown in Ley’s Figure 3, which I have reproduced below, with annotations added to show the ten somatic and

³ The Roche/454 FLX system that used sequencing by synthesis but with a different method of detecting the nucleotides (pyrosequencing) as they were added in sequence as compared to Illumina sequencing. Pyrosequencing relies on light detection based on a chain reaction when pyrophosphate is released.

two germline (inherited) mutations that were analyzed at different stages of the disease (red and green bars, where somatic mutations should primarily be seen), and compared to the skin sample (control, yellow bars, where germline variants should be frequent) that was collected at the same time as the primary tumor.

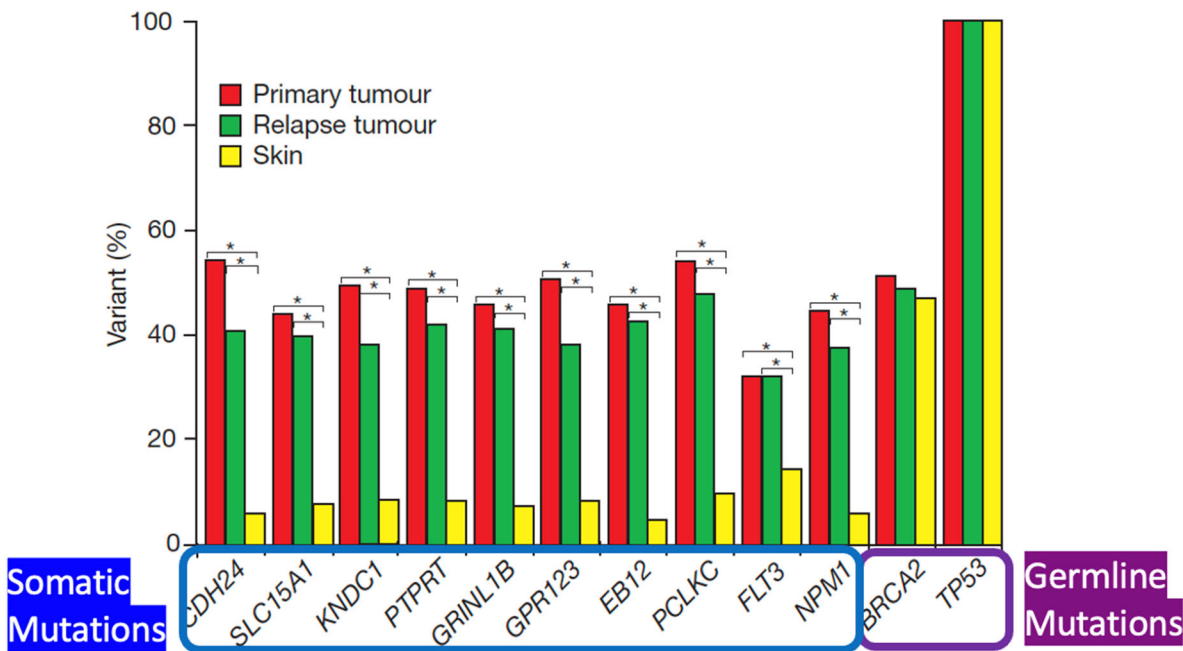


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that of the skin sample for all mutations ($P < 0.000001$ for all mutations, Fisher’s exact test, denoted by a single asterisk in all cases). Note that the normal skin sample was contaminated with leukaemic cells containing the somatic mutations. The patient’s white blood cell count was 105,000 (85% blasts) when the skin punch biopsy was obtained.

84. A key insight of Ley is that characterizing the patient’s tumor using whole-genome sequencing to identify a patient-specific polymorphic fingerprint and then using that data to target specific polymorphisms identified in the whole-genome

sequencing step to track the tumor over time is a useful approach to cancer monitoring and testing. Specifically, Ley observed that “[o]ur results strongly support the notion that hypothesis-driven (for example, candidate gene-based) examination of tumour genomes by PCR-directed or capture-based methods is inherently limited, and will miss key mutations.” *Id.* at 70. A POSA would have appreciated that in Ley’s approach, which uses an initial whole-genome sequencing step to understand the nature of the patient’s tumor, the results of the whole-genome sequencing would help avoid “miss[ing] key mutations,” and only after that information is collected, a specific set of probes for further sequencing of key mutations can be created. *See* Ex. 1031 (Meyerson) at 687 (describing the use of whole-genome sequencing to discover genomic alteration events in cancer tissue in Ley and subsequent studies); 694 (“[C]urrent efforts are focused on applying deep targeted second-generation sequencing to validate the events.”); Fig. 2 (explaining use of hybrid capture for targeted sequencing in cancer genomics applications).

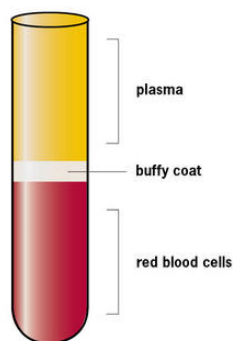
85. Soon after, Leary (Ex. 1002) and McBride (Ex. 1015) demonstrated the use of the same concept, by using whole-genome sequencing to identify structural variants (as opposed to SNVs as in Ley), and to subsequently use those identified markers as targets for serial monitoring of tumor DNA in the blood plasma⁴ of

⁴ Whole blood is often centrifuged to separate it into components or fractions for further analysis. These include plasma, buffy coat, and red blood cells. The

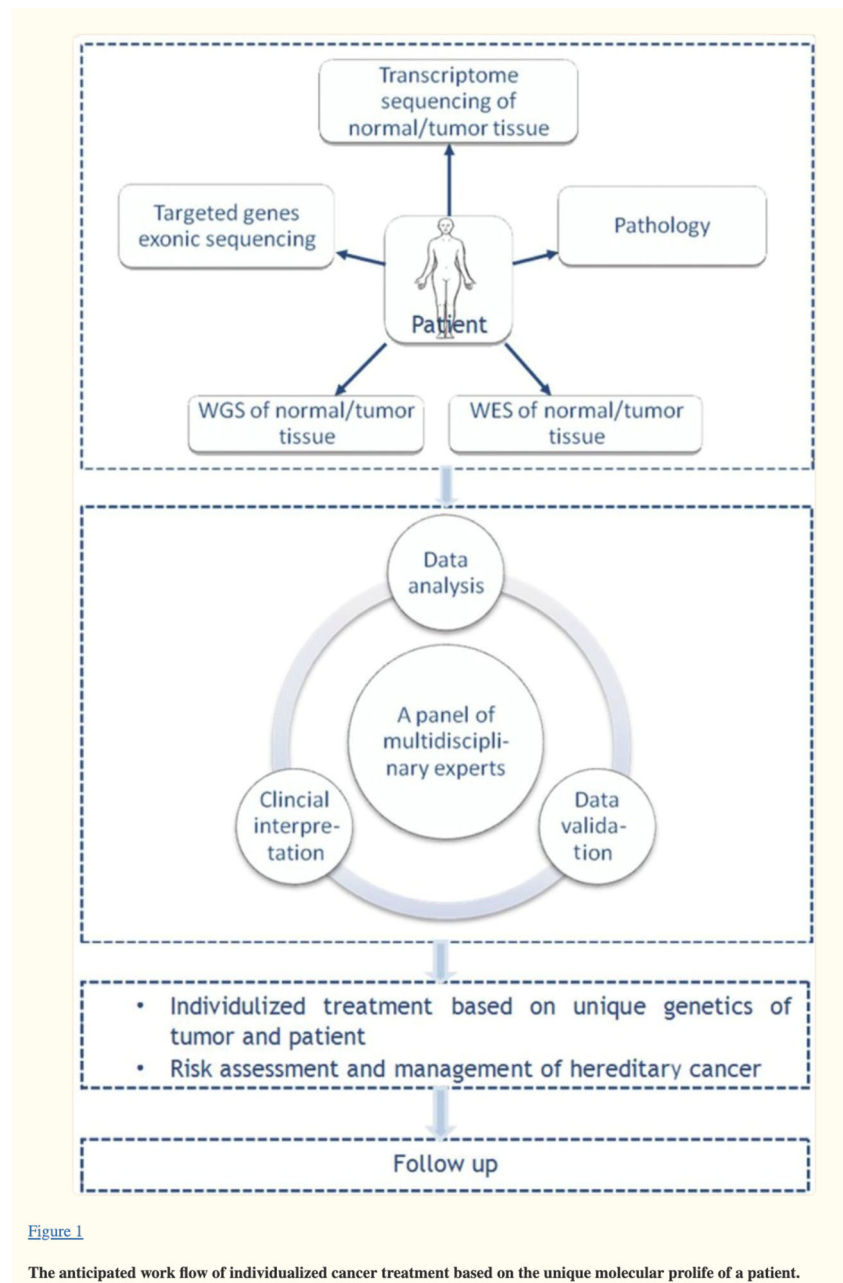
patients with solid tumors. *See also* Ex. 1031 (Meyerson) at 687 (Table 1) (listing Ley and subsequent whole cancer genome studies).

86. By 2012, Guan, which cites Ley’s role in enabling “efficient and accurate detection of novel and rare somatic mutations” in cancer using NGS, surveyed the literature to that point and found that “whole-genome (or whole-exome sequencing” enabled the discovery of “numerous novel genetic aberrations and associated potential therapeutic targets.” Ex. 1106. Guan further observed the important role of NGS “to improve rationally designed individualized medicine,” and proposed a framework by which “patients will have their normal and tumor

plasma is commonly used as a source of cell-free DNA and the buffy coat fraction is an abundant source of leukocytes. *See* Ex. 1001 at 53:60-62 (when a “blood sample is centrifuged to separate the blood,” the resulting fractions are “plasma, buffy coat, and red blood cells. The plasma contains cfDNA and the buffy coat contains leukocytes”). I note that this passage from the specification merely documents the knowledge of a POSA. *See* Ex. 1109 at 29 (“Buffy coat is a leukocyte-enriched fraction of whole blood.” Preparing a buffy-coat fraction from whole blood is simple and yields approximately 5-10 times more DNA than an equivalent volume of whole blood.).



genomes sequenced before doctors design personalized treatment regimens,” because “[w]ith the improvement of NGS technology, it is possible to have a comprehensive molecular profile for a patient available within a clinically acceptable timeframe and cost.” *Id.* Guan further observed that such personalized methods could take advantage of NGS as a “cost-effective method to detect and measure . . . tumor gene mutations in the plasma” by using “mutations in the circulating DNA.” *Id.*



E. Related Applications of Next Generation Sequencing

87. In parallel with the development of sequencing-based techniques for cancer monitoring, those skilled in the art were also applying very similar techniques to the detection and monitoring of other medical conditions. Thus, very similar target selection, target enrichment, and target monitoring techniques were simultaneously

developed in fields such as prenatal genetic testing and organ transplant rejection. This is particularly true in applications that use cell-free DNA (“cfDNA”) from the patient as a marker of disease.

88. Cell-free DNA (“cfDNA”) refers to small fragments of DNA present in the bloodstream. This generally comes from DNA released during apoptosis, or the death of cells, in the body. During apoptosis, cells undergo various cell death processes. Part of that involves cellular mechanisms that cause the DNA of the cells’ chromosomes to be cleaved and fragmented. The fragmented DNA, along with other parts of the cells, are released into the bloodstream, and eventually cleared out of the body with other bodily waste products.

89. Cells are constantly dying and being replaced throughout the body. Therefore, there is always cfDNA being released into and circulating through the bloodstream (or other body fluids, including cerebrospinal fluid and urine) before it is cleared out of the body. Thus, tumors are constantly releasing cfDNA into the patient’s bloodstream, a fetus is constantly releasing cfDNA into its mother’s bloodstream, and a transplanted organ (*e.g.*, a liver or heart transplant) is constantly releasing cfDNA into the transplant recipient’s bloodstream. By identifying distinctions between the “foreign” DNA – whether from a tumor, fetus, or transplanted organ – and the “host” DNA, the levels of foreign DNA present in the host can be used as powerful markers of biomedical information.

90. Before the priority date, those skilled in the art recognized the applicability of similar nucleic acid analysis workflows to each of these unique clinical settings. For example, by 2011, Dr. Stephen Quake had aptly described these connections:

Particularly for cancer, where the circulating DNA bears hallmark signs of the disease including mutations in oncogenes, microsatellite alterations, and, for certain cancers, viral genomic sequences, DNA or RNA in plasma has become increasingly studied as a potential biomarker for disease. For example, Diehl et al recently demonstrated that a quantitative assay for low levels of circulating tumor DNA in total circulating DNA could serve as a better marker for detecting the relapse of colorectal cancer compared with carcinoembryonic antigen, the standard biomarker used clinically. Maheswaran et al reported the use of genotyping of circulating cells in plasma to detect activating mutations in epidermal growth factor receptors in lung cancer patients that would affect drug treatment. *These results collectively establish both circulating DNA, either free in plasma or from circulating cells, as useful species in cancer detection and treatment. Circulating DNA has also been useful in healthy patients for fetal diagnostics, with fetal DNA circulating in maternal blood serving as a marker for gender, rhesus D status, fetal aneuploidy, and sex-linked disorders.* Fan et al recently demonstrated a strategy for detecting fetal aneuploidy by shotgun sequencing of cell-free DNA taken from a maternal blood sample, a methodology that can replace more invasive and risky techniques such as amniocentesis or chorionic villus sampling.

Ex. 1003 (Quake) at ¶40 (citations omitted); *see also id.* at ¶41 (“In all these applications of circulating nucleic acids, the presence of sequences differing from a patient’s normal genotype has been used to detect disease. In cancer, mutations of genes are a telltale sign of the advance of the disease; in fetal diagnostics, the detection of sequences specific to the fetus compared to maternal DNA allows for analysis the health of the fetus.”). Quake then observed that “as cell-free DNA or RNA often arises from apoptotic cells, the relative amount of donor-specific sequences in circulating nucleic acids should provide a predictive measure of on-coming organ failure in transplant patients for many types of solid organ transplantation” *Id.* at ¶42.

F. Providing Actionable Information to Healthcare Providers

91. The ability to monitor and analyze a patient’s cancer using these techniques was an important advance. Those skilled in the art recognized that it was also important to process the obtained sequencing information into a form that would be clinically useful: “Adoption of clinical NGS testing will place significant demands on laboratory infrastructure and will require extensive computational expertise and a deep knowledge of cancer medicine and biology to generate truly useful ‘clinically actionable’ reports.” Ex. 1014 (Ross 2011) at 527.

92. Studies of cancer biology yielded discoveries of associations and relationships between certain genetic mutations and differential effects on cancer,

such as response to treatment, severity of response, and more. For example, as Ross summarizes:

In late 1998, the era of personalized oncology began with the simultaneous regulatory approval of the anti-HER2-targeted monoclonal antibody therapeutic, trastuzumab, for the treatment of HER2-overexpressing breast cancer and the immunohistochemical-based diagnostic HercepTest . . . for the identification of patients who are eligible for treatment. During the ensuing 12 years, interest in these “drug-test” combinations, also referred to as “companion diagnostics,” has increased exponentially as more anticancer agents have been approved and more new drugs have entered clinical trials based on biomarker profiles. *The selection of anti-epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors in non-small cell lung cancer (NSCLC) and the use of anti-EGFR monoclonal antibody therapeutics in colorectal cancer (CRC) are now firmly based on pretreatment sequence determinations for hot spots in the EGFR (NSCLC) and KRAS (CRC) genes.*

Ex. 1014 (Ross 2011) at 527. Ross further notes that more-recently developed therapies based on “sequence detection-based companion diagnostics” had shown impressive results.” *Id.* (discussing PLX-4032 drug candidate targeting cancers “that feature the V600E mutation of the *BRAF* gene,” and crizotinib having “impressive efficacy for the treatment of NSCLCs that feature an *EML4-ALK* translocation”).

93. “Given the potential generation of ‘actionable’ gene sequence abnormalities and abnormalities of unknown clinical importance,” Ross reported, “the successful NGS-based clinical service will have a result report that is focused on the abnormalities that can lead directly to the best available therapy, direct a patient to an ongoing clinical trial, or result in some other actionable response from the treating oncologist.” *Id.* At 537. And these sequencing-based technologies can be used with other forms of cancer cell diagnostics such that these data points can “be combined with tumor cell DNA sequencing in some form of integrated laboratory report.” *Id.* at 536.

94. Such reports were described in a number of prior art references, which depicted information about the relative abundance or a frequency of polymorphic alleles present in a patient’s sample at multiple points in time over the course of treatment. For example, Ley’s Figure 3 depicts sequence information regarding the relative abundance of particular somatic and germline mutations that were observed in samples taken at (i) primary onset, and (ii) relapse. Ex. 1012 (Ley) at 69 (Fig. 3). These abundances are compared to the frequency of the same polymorphic alleles present in skin, which represents a control.

VI. Prior Art References of the Grounds

A. Chan (Ex. 1008)

95. Chan et al., “Cancer Genome Scanning in Plasma: Detection of Tumor-Associated Copy Number Aberrations, Single-Nucleotide Variants, and Tumoral Heterogeneity by Massively Parallel Sequencing,” is an article that was published in *Clinical Chemistry* by January 1, 2013. See Oxford Academic Press Website (<https://academic.oup.com/clinchem/article/59/1/211/5622098>) (“Published 01 January 2013). *Clinical Chemistry* is a well-known and respected journal in the field, and those skilled in the art routinely rely on the publication date information provided by the publishers of major, reputable journals.

B. Okosun-II (Ex. 1102)

96. Okosun et al., “Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma,” is an article published in *Nature Genetics* by February 2014, and which was published online on December 22, 2013. Ex. 1102 at 176. Initial data on the key findings published in this article were published on November 16, 2012 in Okosun et al., “Whole Genome Sequencing in Sequential Biopsies Reveals the Genetic Evolution of Follicular Lymphoma to Transformed Follicular Lymphoma,” which was an abstract in *Blood*, the journal of the American Society of Hematology. Ex.

1101. In the discussion below, I refer to the 2012 abstract (Ex. 1101) as Okosun-I, and the 2013 article (Ex. 1102) as Okosun-II.

C. Ley (Ex. 1012)

97. Ley *et al.*, “DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome,” is an article that was published in *Nature* by November 6, 2008, and is thus prior art to the ‘968 patent. Ley “establishes whole-genome sequencing as an unbiased method for discovering cancer-initiating mutations in previously unidentified genes that may respond to targeted therapies.” Ex. 1012 at 66. See Section ¶V.D, above.

VII. Claim Construction

98. I have generally applied the plain and ordinary meaning that the claims of the ‘968 patent would have had to a POSA in view of the ‘968 patent’s specification, and as Patent Owner has read the claims in its complaint for infringement against Petitioner. Below I set forth my understanding of several claim terms at issue, which I have applied in each of the grounds of invalidity set forth below.

A. “A method of analyzing a biological sample obtained from a subject being screened for a cancer” – Claim 1

99. I have been informed by counsel that whether the preamble of a claim – here “[a] method of analyzing a biological sample obtained from a subject being screened for a cancer” – is considered to be a claim limitation is a question of law,

and that there are situations in which a preamble is not a claim limitation. Nevertheless, in my analysis below, I have assumed that the preamble is a limitation of the claims, and have identified its presence in the cited anticipatory prior art and obviousness combinations.

100. The ‘968 patent notes that a subject “being screened for a cancer” includes patients who have already been diagnosed with cancer. Ex. 1001 at 1:19-23 (“[I]t can be difficult to monitor the progression of disease after it has been detected, especially if it requires screens which are expensive and/or invasive.”).

B. “mosaic variants” – Claim 1

101. I have applied the understanding of “mosaic variants” that Patent Owner has applied in its First Amended Complaint for infringement against Petitioner. Specifically, Patent Owner stated:

As those in the art understand, somatic cancer-causing variations are a type of mosaic variant. This is reflected for instance, in a publication by Wilson et al.:

The most well-studied example of somatic mosaicism is cancer; all tumors represent a subset of cells within an individual that have accumulated mutations resulting in uncontrolled cell growth and survival advantage (16). Both benign and malignant tumors constitute evidence of how common somatic mosaicism is in the human body. . . .

Furthermore, NGS is also useful in identifying mutations

underlying cancers (78). ***Cancer is a condition of mosaicism in two folds***: first, mosaicism is achieved by accumulation of mutations that confer survival advantage to the mutant cells and give rise to tumors (16); second, each mutant cell accumulates additional genetic alterations, due to lack of cell cycle checkpoints and DNA repair mechanisms, creating mosaicism within tumors (tumor heterogeneity) (79). Multiple NGS studies have successfully identified these two types of mosaicism, leading to a better understanding of different molecular subtypes of cancer and allowing for targeted therapy for each specific tumor and/or patient (80).

Cohen Wilson and Trinh Ye, *Detecting Somatic Mosaicism: Considerations and Clinical Implications*, 87 CLINICAL GENETICS 554, at 555, 558 (2015).

Ex. 1107at ¶43 (emphasis original). Patent Owner went on to contend that “[t]he patents,” including the ‘968 patent, “are consistent with this understanding and likewise **make clear that somatic cancer-causing variations are a type of mosaic variant.**” *Id.* ¶44.

102. Claim 5 of the ‘968 patent states that “said mosaic variants” of claim 1 are “causal or predisposing for a cancer,” which aligns with the definition advanced by Patent Owner in the sections of its complaint against Petitioner cited above. Claim 6 of the ‘968 patent states that “said mosaic variants” of claim 2 are “non-causal for a cancer.” Thus, under Patent Owner’s understanding of the claims, the claimed

“mosaic variants” include somatic mutations that are causal for a cancer as well as somatic mutations that are non-causal for a cancer. Accordingly, Patent Owner’s understanding of the claims advanced in its infringement complaint, when taken in light of its own dependent claims, requires that the term “mosaic variants” encompasses both somatic cancer-causing variations and somatic non-causal cancer mutations. *See also* Ex. 1108 at 18, 34, 137 (“[T]he claimed methods are not restricted to the identification of cancer driver genes. As discussed in claims 16 and 17, the methods of the claims allow for the detecting of mosaic variants that are causal or predisposing of cancer, as well as mosaic variants that are not causal for cancer.”).

103. Thus, Patent Owner’s understanding of the term “mosaic variants,” requires the claimed “mosaic variants” to encompass any somatic mutation present in a cancer nucleic acid sample (i.e., either causal or non-causal somatic mutations, or both) that is not present in a normal nucleic acid. I have applied this understanding in my analysis below.

C. “capture probe” – Claim 13

104. I have applied the meaning the term “capture probe” would have had to a POSA in view of the ‘968 patent’s specification. The ‘968 patent explains that the main feature of a “capture probe” is that it “typically . . . comprises a nucleic acid binding site.” Ex. 1001 at 35:30-37.

105. The nucleic acid binding sites may comprise a sequence “that hybridizes to at least a portion of the one or more nucleic acid molecules . . . in the sample or subset of nucleic acid molecules.” *Id.* at 35:49-53. Thus, the ‘968 patent explains, “capture probes may comprise a nucleic acid binding site that hybridizes to one or more genomic regions.” *Id.* at 35:53-54.

106. Additional features for capture probes, such as linkers and labels, are described as optional. *See id.* at 35:32-37 (“The capture probe may further comprise one or more labels. The capture probes may further comprise one or more linkers. The one or more linkers may attach the one or more labels to the nucleic acid binding site.”).

107. Accordingly, a POSA would have understood the term “capture probe” in the ‘968 patent to encompass at least capture probes that are PCR primers (which for shorthand I will refer to here as “PCR primer capture probes”) as well as capture probes that are solution-based hybridization baits (e.g., biotinylated oligonucleotide baits) (which for shorthand I will refer to here as “solution-based hybrid capture probes.”). I understand that in related proceedings involving Patent Owner’s ‘394 and ‘033 patents, that the Board adopted this construction for the term “capture probes.” I note that the passage in the specification defining “capture probes” in the ‘968 patent appears to be copied and pasted directly from the specification of those patents, supporting construing the term in the same way here.

VIII. Prior Art Anticipated or Would Have Rendered Obvious Claims 1-13 of the '968 Patent

A. Ground 1: Chan Anticipates Claims 1-4, 7, and 9-12

1. Claim 1

108. Claim 1 recites “A method of analyzing a biological sample from a subject being screened for cancer, wherein the biological sample is from a non-fetal source, the method comprising” enumerated steps a-d. As discussed in detail below, Chan discloses such a method, and thus anticipates claim 1.

109. In full, Claim 1 recites:

A method of analyzing a biological sample obtained from a subject being screened for a cancer, wherein the biological sample is from a non-fetal source, the method comprising:

- (a) obtaining a first set of sequence reads and a second set of sequence reads by sequencing nucleic acid molecules derived from a first set of nucleic acid molecules and a second set of nucleic acid molecules, wherein said first set of sequence reads corresponds to said first set of nucleic acid molecules and said second set of sequence reads corresponds to said second set of nucleic acid molecules, wherein said first set of sequence reads and said second set of sequence reads are generated using whole genome sequencing, and wherein (i) said first set of nucleic acid molecules is extracted from tissue sample of said subject, and (ii) said second set of nucleic acid molecules is extracted from leukocytes from a blood sample;
- (b) using said first set of sequence reads and said second set of sequence reads to identify mosaic variants specific to said tissue sample of said

subject in said first set of sequence reads that is not present in said second set of sequence reads;

(c) subsequent to (b), identifying said mosaic variants specific to said tissue sample of said subject in a first additional sample of said subject, wherein said first additional sample is obtained from a source that is different than said tissue sample and at a time point that is different than said tissue sample; and

(d) providing a report based on an identification of said mosaic variants specific to said tissue sample of said subject in said first additional sample, wherein said report comprises predicting, diagnosing, and/or prognosing a status or outcome of a cancer in said subject.

110. The preamble of claim 1 recites “[a] method of analyzing a biological sample from a subject being screened for a cancer, wherein the biological sample is from a non-fetal source.”

111. Chan discloses such a method – it is titled “Cancer Genome Scanning in Plasma: Detection of Tumor-Associated Copy Number Aberrations, Single-Nucleotide Variants, and Tumoral Heterogeneity by Massively Parallel Sequencing.” Ex. 1008 at 211. Chan recruited human subjects for this cancer scanning (i.e., screening multiple times): “Four hepatocellular carcinoma patients and a patient with synchronous breast and ovarian cancers were recruited. DNA was extracted from the tumor tissues, and the preoperative and postoperative plasma samples of these patients were analyzed with shotgun MPS.” *Id.* Thus, Chan’s

screening involves the analysis of samples from non-fetal sources. *See also id.* at 212 (explaining close relationship of the disclosed cancer screening methods to the authors' experience in "fetal genomic scanning").

112. The method disclosed by Chan meets each limitation of claim 1, as discussed below.

a) [1] obtaining a first set of sequence reads and a second set of sequence reads by sequencing nucleic acid molecules derived from a first set of nucleic acid molecules and a second set of nucleic acid molecules, [2] wherein said first set of sequence reads corresponds to said first set of nucleic acid molecules and said second set of sequence reads corresponds to said second set of nucleic acid molecules, [3] wherein said first set of sequence reads and said second set of sequence reads are generated using whole genome sequencing, and [4] wherein (i) said first set of nucleic acid molecules is extracted from tissue sample of said subject, and (ii) said second set of nucleic acid molecules is extracted from leukocytes from a blood sample

113. In sum, step 1(a) requires performing whole-genome sequencing (WGS) on nucleic acid molecules extracted from (i) a tissue sample from the subject; and (ii) leukocytes from a blood sample.⁵ Chan discloses step 1(a) by performing WGS on matched (i) tumor tissue from hepatocellular carcinoma (HCC); and (ii)

⁵ The claim does not specify the source of the blood sample (i.e., unlike claim element 1(a)(i), it does not specify that the leukocytes are "from a blood sample of *said subject*"). Nevertheless, Chan discloses obtaining leukocytes from a blood sample of said subject.

buffy coat DNA, which is the fraction of a centrifuged blood sample that contains leukocytes.⁶ Ex. 1008 at 211 (“Four hepatocellular carcinoma patients . . . were recruited. DNA was extracted from tumor tissues, and the preoperative and postoperative plasma samples of these patients were analyzed with shotgun MPS.”); *id.* at 218 (Describing identification of mosaic variants by comparing tumor and germline DNA using WGS: “The MPS data from the tumor DNA and the buffy coat DNA from each of the 4 HCC patients were compared, and SNVs present in the tumor DNA but not in the buffy coat DNA were mined with a stringent bioinformatics algorithm.”).

114. Given the prolixity of the claim language, I have broken step 1(a) into bracketed subparts [1]-[4] in the analysis that follows.

(1) Step 1(a)([1]): “obtaining a first set of sequence reads and a second set of sequence reads by sequencing nucleic acid molecules derived from a first set of

⁶ See Ex. 1001 at 53:60-62 (when a “blood sample is centrifuged to separate the blood,” the resulting fractions are “plasma, buffy coat, and red blood cells. The plasma contains cfDNA and the buffy coat contains leukocytes . . .”). I note that this passage from the specification merely documents the knowledge of a POSA. See Ex. 1109 at 29 (“Buffy coat is a leukocyte-enriched fraction of whole blood. Preparing a buffy-coat fraction from whole blood is simple and yields approximately 5-10 times more DNA than an equivalent volume of whole blood.”).

nucleic acid molecules and a second set of nucleic acid molecules.”

115. Chan discloses step 1(a)([1]): “We sequenced tumor DNA and buffy coat DNA to mean depths of 29.5-fold (range, 27-fold to 33-fold) and 43-fold (range, 39-fold to 46-fold) haploid genome coverage, respectively.” Ex. 1008 at 218.

116. Specifically, Chan discloses “obtaining a first set of sequence reads . . . by sequencing nucleic acid molecules derived from a first set of nucleic acid molecules” because it sequences, and thus obtains sequence reads for, tumor DNA. *Id.*; *see also* Table 2. In parallel, Chan discloses “obtaining . . . a second set of sequencing reads by sequencing nucleic acid molecules derived from . . . a second set of nucleic acid molecules,” because it sequences, and thus obtains sequence reads for, “buffy coat DNA.” *Id.* at 218; *see also* Table 2.

(2) Step 1(a)([2]): “wherein said first set of sequence reads corresponds to said first set of nucleic acid molecules and said second set of sequence reads corresponds to said second set of nucleic acid molecules”

117. The sequence reads obtained in Step 1(a)([1]) discussed above were necessarily obtained by sequencing DNA molecules, *see* Section V.C, and the sequence reads necessarily correspond to the sample from which they were obtained. In other words, the first set of sequence reads corresponds to the nucleic acid sequence (i.e., A, C, G, T) of the tumor DNA molecules and the second set of

sequence reads corresponds to the nucleic acid sequence of the buffy coat DNA molecules.

(3) Step 1(a)([3]): “wherein said first set of sequence reads and said second set of sequence reads are generated using whole genome sequencing,”

118. Chan discloses that the first and second sets of sequence reads are generated using “MPS” or “massively parallel sequencing,” and that the “tumor DNA and buffy coat DNA” was sequenced “to mean depths of 29.5-fold . . . and 43-fold . . . haploid genome coverage, respectively.” Ex. 1008 at 218. This was non-targeted whole-genome sequencing to “explore[] the *genomewide* detection of tumor-derived SNVs in the plasma of the 4 HCC patients.” *Id.* To inform subsequent genomewide scanning in the plasma, Chan first subjected the HCC patients’ tumors to an initial paired tumor/normal sequencing step, in which a genome-wide comparison identified “SNVs present in the tumor DNA but not in the buffy coat DNA” based on sequencing to >29-fold haploid coverage of each sample. *Id.* at 218. Thus, Chan’s first and second sets of sequence reads are generated using whole genome sequencing.

119. Chan’s materials and methods section further confirms that the “MPS” of “tumor DNA and buffy coat DNA” to depths of 29.5-fold and 43-fold haploid genome coverage was “whole genome sequencing.” *Id.* at 212 (“DNA Sequencing”). Chan reported that for DNA obtained from sources other than plasma samples, i.e.,

the tumor and buffy coat DNA samples, “Sequencing libraries of the genomic DNA samples were constructed with the Paired-End Sample Preparation Kit (Illumina),” which involved shearing genomic DNA into 200-bp fragments. *Id.* The prepared DNA libraries were then sequenced “by 51 x 2 cycles or 76 x 2 cycles of sequencing on a HiSeq 2000 system (Illumina) with the TruSeq SBS Kit (v2).”

120. In IPR proceedings involving related Personalis patents, I understand that Personalis has taken the position that “whole genome sequencing” requires “a non-targeted approach to sequencing all or almost all of the genome.” Ex 1116 (394 POR) at 5-6. If that construction is applied, Chan’s MPS of the tumor and buffy coat samples satisfies the definition. First, Chan’s approach is “non-targeted” because the sequencing libraries are prepared from randomly sheared genomic DNA, and thus reads are intended to be generated from all DNA in a sample, as opposed to targeted sequencing of chosen or selected DNA loci (e.g., the exome only). Second, Chan’s approach sequences all or almost all of the genome, specifically noting “haploid genome coverage” of 29.5-fold (tumor DNA) and 43-fold (buffy coat DNA). Thus, on average, each base of the haploid genome was sequenced 29.5 times in the tumor sample, and 43 times in the buffy coat sample, confirming that Chan’s nontargeted sequencing approach sequenced all or almost all of the genome.

(4) Step 1(a)([4]) “wherein (i) said first set of nucleic acid molecules is extracted from tissue sample of said subject, and (ii) said second set of nucleic acid

molecules is extracted from leukocytes from a blood sample”

121. As explained above, Chan performed whole-genome sequencing on (i) “DNA [] extracted from the tumor tissues” of HCC patients and (ii) buffy coat DNA, which is DNA from the fraction of a blood sample that contains leukocytes. Ex. 1008 at 211, 218. *See* Ex. 1001 at 53:60-62 (when a “blood sample is centrifuged to separate the blood,” the resulting fractions are “plasma, buffy coat, and red blood cells. The plasma contains cfDNA and *the buffy coat contains leukocytes . . .*”). *See also id.* at 212 (“Genomic DNA Extraction[:] Genomic DNA was extracted from the patients’ buffy coat samples according to the blood and body fluid protocol of the QIAmp DSP DNA Blood Mini Kit. DNA was extracted from tumor tissues with the QIAmp DNA Mini Kit (Qiagen).”).

b) Step 1(b): “using said first set of sequence reads and said second set of sequence reads to identify mosaic variants specific to said tissue sample of said subject in said first set of sequence reads that is not present in said second set of sequence reads;”

122. After obtaining whole-genome sequencing reads, Chan discloses step 1(b) because it performs a bioinformatics analysis that identifies mosaic variants specific to the tissue (tumor) sample by comparing SNVs in tumor and normal tissue to identify such variants that are unique to the tumor, and thus not present in the germline sequence reads. Ex. 1008 at 211, 218. Chan explains that “[t]he MPS [i.e.,

WGS] data from the tumor DNA and the buffy coat DNA from each of the 4 HCC patients were compared, and *SNVs present in the tumor DNA but not in the buffy coat DNA were mined* with a stringent bioinformatics algorithm.” *Id.* at 218. Chan explained that “[t]he number of tumor-associated SNVs,” which constitute mosaic variants under Patent Owner’s understanding of the claim, “ranged from 1334 to 3171 in the 4 HCC cases.” *Id.*

c) Step 1(c): “subsequent to (b), identifying said mosaic variants specific to said tissue sample of said subject in a first additional sample of said subject, wherein said first additional sample is obtained from a source that is different than said tissue sample and at a time point that is different than said tissue sample;”

123. After identifying mosaic variants – tumor specific SNVs – in the HCC patients based on the paired tumor/germline whole-genome sequencing assay, Chan discloses Step 1(c) by identifying at least some of the previously identified mosaic variants in additional samples. Those samples include (i) a preoperative plasma sample; and (ii) a post-operative plasma sample. Ex. 1008 at 211, 215, 218. Both such samples constitute additional samples that are obtained from a source that is different from the tumor tissue sample (at a minimum because such samples are taken at distinct time points from blood plasma). And both such samples were taken at a time point that is different than said tissue sample (which was taken during surgery): the preoperative plasma samples were taken “before . . . surgical resection

with curative intent from [the] 4 HCC patents,” and the postoperative plasma samples were taken “1 week after surgical resection with curative intent from [the] 4 HCC patients.” Ex. 1008 at 215.

124. After identifying the mosaic variants – in the form of tumor specific SNVs – Chan then evaluated “[t]he proportions of such SNVs that were detectable in plasma,” and listed them in Table 2. *Id.* at 218. “Before treatment, 15%-94% of the tumor-associated SNVs were detected in plasma.” *Id.* at 218. After treatment, the mosaic variants – tumor specific SNVs – were detected in plasma as at a rate of 1%-5% (1.5%, 2.3%, 1%, and 5.5% in patients HCC1-HCC4, respectively). *Id.* at 218 (Table 2) (reproduced below); *id.* at 219 (“These fractional concentrations . . . were reduced after surgery (Table 2).”).

Table 2. Fractional concentrations of tumor-derived DNA in plasma determined by SNV analysis.							
Case	No. of SNVs detected in tumor tissue	Time point	No. of tumor-associated SNVs sequenced from plasma (percentage of SNVs seen in tumor tissue)	No. of plasma DNA sequence reads showing SNVs (<i>p</i>)	No. of plasma DNA sequence reads showing wild-type sequence (<i>q</i>)	Deduced fractional concentration of tumor-derived DNA by SNV analysis $\left(\frac{2p}{p+q}\right)$	Deduced fractional concentration of tumor-derived DNA by GAAL analysis
HCC1	2840	Before surgery	2569 (94%)	11 389	31 602	53%	52%
		After surgery	44 (1.5%)	91	46 898	0.4%	0.9%
HCC2	3105	Before surgery	1097 (35%)	1490	57 865	5.0%	5.6%
		After surgery	72 (2.3%)	206	66 692	0.6%	0.9%
HCC3	3171	Before surgery	461 (15%)	525	48 886	2.1%	4.3%
		After surgery	31 (1%)	67	58 862	0.2%	1.4%
HCC4	1334	Before surgery	201 (15%)	248	18 527	2.6%	7.6%
		After surgery	74 (5.5%)	149	22 144	1.3%	2.7%

d) Step 1(d): providing a report based on an identification of said mosaic variants specific to said tissue sample of said subject in said first additional sample, wherein said report comprises predicting, diagnosing, and/or prognosing a status or outcome of a cancer in said subject.

125. Chan discloses Step 1(d) by providing a report based on the identification of mosaic variants – tumor specific SNVs – in preoperative and postoperative plasma samples. Ex. 1008 at 218 (Table 2). This report shows the “Fractional concentrations of tumor-derived DNA in plasma determined by SNV [mosaic variant] analysis,” and shows that for each of the HCC patients analyzed, the number of tumor-associated SNVs sequenced from plasma decreased following “surgical resection with curative intent” as compared to presurgical plasma samples. *Id.* at 215, 218. The decrease in fraction of tumor specific SNVs (i.e., the mosaic variants) present after surgery constitutes a report that predicts, diagnoses, and/or prognoses the status or outcome of the patient’s cancer, namely whether the surgery increased or decreased the level of HCC biomarkers present in the patient. Here, the decrease in detected mosaic variants (tumor specific SNVs) in the post-operative plasma indicates that the surgery was successful at least in reducing such biomarkers. *See id.* at 221 (“This approach would allow qualitative and quantitative aspects of such aberrations to be followed serially, as evidenced by the changes we observed in the pre- and postsurgery plasma samples. The method would also allow tumor load to be assessed, as indicated by the correlation between the fractional

tumor DNA concentration measured and the size of the tumor. In this study, these general characteristics were observed in the plasma DNA profiles derived from HCC, breast cancer, and ovarian cancer.”). *See* Ex. 1001 at 14:19-34 (“The methods may comprise reporting the number of reads, localization, cell type, spatial position in the body, and/or other values based on the comparison. . . The methods may comprise reporting the changes in the overall set of sequences.”).

126. Thus, Chan anticipates claim 1.

2. Claim 2

127. Claim 2 recites, “[t]he method of claim 1, further comprising identifying mosaic variants specific to said tissue sample of said subject in a second additional sample of said subject, wherein said second additional sample comprises cell-free DNA, which second additional sample is obtained at a time point that is different than said tissue sample and said first additional sample.”

128. As explained with respect to Step 1(c) above, Chan discloses identifying tumor specific SNVs (mosaic variants specific to the HCC tissue sample) of each patient in two additional plasma samples, obtained pre- and post-operatively. Thus, Chan analyzes at least three samples at three time points: (1) the pre-operative plasma samples at a time point before surgery; (2) the surgically resected tumor sample at the time of the surgery; and (3) the post-operative plasma sample one week

after surgery. Ex. 1008 at 215, 218. Chan’s plasma samples alternatively satisfy the claimed first and second additional samples (i.e., the preoperative plasma sample can constitute the first additional sample, in which case the postoperative plasma sample constitutes the second additional sample, and vice versa).

129. Moreover, Chan’s plasma samples comprise cell-free DNA. *See* Ex. 1008 at 212 (“Cell-free DNA molecules from 4.8 mL of plasma were extracted . . . for subsequent preparation of the DNA-sequencing library.”). *See* also Ex. 1001 at 53:62 (“The plasma contains cfDNA.”⁷).

130. Thus, Chan anticipates claim 2.

3. Claim 3

131. Claim 3 recites, “[t]he method of claim 2, wherein said first and/or second additional sample comprises a blood sample, a saliva sample, a buccal swab, a urine sample, a semen sample, or a cerebrospinal fluid sample of said subject.”

132. As explained with respect to claim 2, above, the first and second additional samples disclosed by Chan are plasma samples, which are extracted from blood samples. The ‘968 patent acknowledges that a blood sample may consist of extracted fractions (e.g., plasma, buffy coat, red blood cells) of the blood sample. Ex. 1001 at 3:12-13 (“In some embodiments, the blood sample is a whole blood

⁷ This statement reflects the knowledge of a POSA, not an allegedly inventive contribution of the ‘968 patent.

sample.”); 45:35-42; (“In some cases, the sample is a blood sample. The sample may comprise one or more peripheral blood lymphocytes. The sample may be a whole blood sample. The blood sample may be a peripheral blood sample. In some cases, the sample comprises peripheral blood mononuclear cells (PBMCs); in some cases, the sample comprises peripheral blood lymphocytes (PBLs). The sample may be a serum sample.”); 53:59-64 (“A blood sample is taken from a patient during a routine visit to the clinic. The blood sample is centrifuged to separate the blood into plasma, buffy coat, and red blood cells.”).

133. Thus, Chan anticipates claim 3.

4. Claim 4

134. Claim 4 recites, “[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are selected from the group consisting of: copy number variation, small insertions and deletions (inDels), single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), and any combination thereof.”

135. As explained above with respect to claims 1 and 2, Chan discloses identifying mosaic variants specific to the tumor tissue sample in the form of tumor specific single nucleotide variants, which are also known as single nucleotide polymorphisms.

136. Thus, Chan anticipates claim 4.

5. Claim 7

137. Claim 7 recites, “[t]he method of claim 2, further comprising enriching said extracted nucleic acid molecules of said tissue sample, blood sample, first additional sample, or second additional sample.”

138. Chan discloses enriching the extracted DNA samples used in its WGS assay (tissue sample and buffy coat blood sample) in the preparation of sequencing libraries: “The adapter-ligated DNA was enriched with 12-cycle PCR.” Ex. 1008 at 2112. Chan also discloses enriching the plasma samples used in the first and second additional samples: “Because the plasma DNA molecules were short fragments [(21)] and the amounts of total DNA in the plasma samples were relatively small, we omitted the fragmentation steps and used a 15-cycle PCR when constructing the DNA libraries from the plasma samples.” *Id.*

139. Thus, Chan anticipates claim 7.

6. Claim 9

140. Claim 9 recites, “[t]he method of claim 1, wherein said tissue sample comprises a metastatic tissue sample, a mixed tissue sample, a biopsy, or a core needle biopsy.”

141. As explained above, the tumor tissue sample was surgically resected, and thus constitutes a biopsy. *See* Ex. 1008 at 212. Moreover, given that a surgically

resected HCC tumor sample will inevitably and necessarily contain some normal liver cells, such samples also constitute a mixed tissue sample.

142. Thus, Chan anticipates Claim 9.

7. Claim 10

143. Claim 10 recites, “[t]he method of claim 1, further comprising enriching said extracted nucleic acid molecules of said tissue sample, blood sample, or first additional sample.”

144. Chan discloses enriching at the extracted DNA samples used in its WGS assay (tissue sample and buffy coat blood sample) in the preparation of sequencing libraries: “The adapter-ligated DNA was enriched with 12-cycle PCR.” Ex. 1008 at 2112. Chan also discloses enriching the plasma samples used in the first and second additional samples: “Because the plasma DNA molecules were short fragments [(21)] and the amounts of total DNA in the plasma samples were relatively small, we omitted the fragmentation steps and used a 15-cycle PCR when constructing the DNA libraries from the plasma samples.” *Id.*

145. Thus, Chan anticipates claim 10.

8. Claim 11

146. Claim 11 recites, “[t]he method of claim 1, wherein said mosaic variants specific to said subject are selected from the group consisting of: copy number variation, small insertions and deletions (inDels), single nucleotide

polymorphisms (SNPs), multiple nucleotide polymorphisms ([M]NPs), and any combination thereof.”

147. As explained above with respect to claim 1, Chan discloses identifying mosaic variants to the tumor tissue sample in the form of tumor specific single nucleotide variants, which are also known as single nucleotide polymorphisms.

148. Thus, Chan anticipates claim 11.

9. Claim 12

149. Claim 12 recites, “[t]he method of claim 1, wherein said predicting, diagnosing, and/or prognosing of (d) comprises one or more of the following: (i) diagnosing a cancer, (ii) identifying a cancer, (iii) determining a stage of a cancer, (iv) assessing risk of a cancer, (v) assessing risk of cancer recurrence, (vi) assessing reproductive risk, (vii) assessing efficacy of a drug, (viii) assessing risk of an adverse drug reaction, (ix) predicting optimal drug dosage, (x) predicting drug resistance, or any combination thereof.”

150. Chan’s predicting, diagnosing, and/or prognosing at a minimum comprises (i) diagnosing the patient’s HCC cancer as HCC and (ii) identifying the patient’s HCC cancer as HCC because the plasma samples it sequences are analyzed to determine their relationship to SNVs detected in tumor tissue. *See* Ex. 1008 at 218, Table 2.

151. Thus, Chan anticipates claim 12.

B. Ground 2: Claims 1-13 Would Have Been Obvious over Chan in view of the Knowledge of a POSA

152. As discussed above, Chan anticipates independent claims 1 and 2. Chan renders the challenged claims obvious for at least the following additional reasons.

1. Claim 2

153. Claim 2 recites, “[t]he method of claim 1, further comprising identifying mosaic variants specific to said tissue sample of said subject in a second additional sample of said subject, wherein said second additional sample comprises cell-free DNA, which second additional sample is obtained at a time point that is different than said tissue sample and said first additional sample.”

154. As explained with respect to Step 1(c) above, Chan discloses identifying tumor specific SNVs (mosaic variants specific to the HCC tissue sample) of each patient in two additional plasma samples, obtained pre- and post-operatively, thus anticipating claim 2. Chan further renders obvious claim 2 because it suggests that a POSA would have been motivated to serially obtain additional plasma DNA samples in order to scan the patient’s cancer genome over time. Ex. 1008 at 221 (“This approach would allow qualitative and quantitative aspects of such aberrations to be followed serially, as evidenced by the changes we observed in the pre- and postsurgery plasma samples.”). Thus, Chan itself recommends analyzing multiple plasma samples containing cell-free DNA in order to serially monitor tumor-specific

mosaic variants. A POSA would have reasonably expected success in performing such serial analysis, given Chan's demonstrated prior success in performing the same analysis on the pre- and post-operative plasma samples.

155. Thus, Chan further renders claim 2 obvious.

2. Claims 5 and 6

156. Claim 5 recites, "[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are causal or predisposing for a cancer."

157. Claim 6 recites, "[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are non-causal for a cancer."

158. The '968 patent does not disclose any information regarding the identification of a particular mosaic variant as "causal" or "non-causal," beyond the tautology that a mosaic variant may be either causal or non-causal. *See, e.g.*, Ex. 1001 at 12:15-17 ("The specific mosaic variants can be used as biomarkers . . . even if they are not causal for medical condition."); *id.* at 13:38-40 ("In some cases, the tumor variants are causal variants (i.e., tumor-causing)."). The specification does not provide any criteria for when a variant should be considered "causal" or not.

159. At a minimum, the mosaic variants identified by Chan necessarily constitute either causal or non-causal mutations. Chan explained that "[i]n this work, we used tumor-associated SNVs as tumor markers and did not focus on elucidating the biology of the mutational landscape of the studied cancers" because its focus was

“to develop a noninvasive approach to assess the tumor burden of the body.” Ex. 1008 at 222. But Chan goes on to explain why a POSA would have found it obvious to do so, and that a POSA would have reasonably expected success in doing so: “We expect, however, that most of the tumor-associated SNVs we detected were ‘passenger’ rather than ‘driver’ mutations.” *Id.* Thus, the SNVs included a mixture of non-causal (passenger) and causal (driver) mutations. Chan explained that a POSA would have wanted to identify passenger mutations because they “provide[] a numerical advantage that has the potential for more sensitive, precise, and representative detection of tumor DNA in plasma,” aiding tumor burden assessments. *Id.* Chan also explains that “selective analyses of driver mutations may allow more readily the identification of medically actionable targets.” *Id.*

160. Accordingly, Chan renders obvious detecting both causal and non-causal mosaic variants, and thus claims 5 and 6.

3. Claim 8

161. Claim 8 recites, “[t]he method of claim 1, further comprising providing a recommendation for a therapeutic intervention based on an identification of said mosaic variants specific to said tissue sample of said subject.”

162. As noted above, Chan explained that “selective analyses of driver mutations may allow more readily the identification of medically actionable targets.” Ex. 1008 at 222. A POSA would have understood that a “medically actionable

target” is a type of causal mosaic variant for which a targeted or companion therapy existed, and thus would be obvious to administer. Moreover, Chan’s analysis of post-surgical plasma mosaic variant levels would have rendered obvious therapeutic intervention, such as follow-up surgery, in the event that such levels indicated a rising tumor burden.

163. Thus, Chan renders claim 8 obvious.

4. Claim 13

164. Claim 13 recites, “[t]he method of claim 1, wherein said mosaic variants specific to said tissue sample of said subject are identified using capture probes in said first additional sample of step (c).”

165. As discussed with respect to claim 1, Chan expressly discloses identifying tumor-specific mosaic variants in a plasma sample of step (c) using whole-genome sequencing. Chan further renders obvious the use of capture probes to identify the tumor-specific mosaic variants in an additional plasma sample, such as a post-operative plasma sample. Specifically, Chan explains that “[o]ne can also envision,” after initial analysis of shotgun whole-genome sequencing data to identify mosaic variants, “targeting such sequences for serial monitoring purposes.” In order to target such sequences for serial monitoring purposes, Chan proposes the use of “solution-phase hybridization-based capture approaches,” which had “already been successfully used to enrich plasma DNA for noninvasive prenatal diagnoses.” *Id.*

223. In light of this experience, a POSA would have reasonably expected success in the use of similar capture techniques to enrich cancer-specific mosaic variants from plasma DNA. *Id.* (“[W]e see that the targeted and shotgun approaches can be used in combination for cancer detection and monitoring.”).

166. Thus, Chan renders claim 13 obvious.

C. Ground 3: Claims 1-13 Would Have Been Obvious over Ley in view of Knowledge of a POSA

167. Ley in view of a POSA’s knowledge would have rendered the challenged claims obvious. As interpreted by Patent Owner, the claims of the ‘968 patent require performing a whole-genome sequencing assay on sources of tumor and germline DNA, identifying “mosaic variants” specific to the tumor DNA, identifying such variants in a later sample, and providing a report on the results of such assays. Such methods had been practiced repeatedly before the November 2014 priority date, beginning no later than Ley’s 2008 disclosure (Ex. 1012).

1. Overview of Ley

168. Because Ley analyzed AML, a type of leukemia, the “tumor sample” was derived from bone marrow or blood, rather than tissue as in the claims. For the germline reference sample, Ley used skin (a tissue biopsy), given that blood cells would serve as a suboptimal source of germline DNA in a leukemia patient because leukemia is a cancer of the blood cells. For example, Ley observed contamination of

its skin sample due to the presence of leukemic blood cells in the tissue biopsy. Ex. 1012 at 69 (Figure 3). Ley confirmed that the skin cells were an appropriate control because “the variant alleles also were detected at frequencies of ~5-13% in the skin sample,” because it was “clear that the skin sample contained contaminating leukaemic cells” due to a high white blood cell count, and thus it was straightforward to remove false positive reads from the skin sample in the discovery pipeline. *Id.* at 69.

169. Example 4 of the ‘968 patent confirms the foundational nature of Ley’s study, as it purports to describe as inventive an experiment that is remarkably similar to what Ley disclosed six years prior:

In this example, a tumor versus normal analysis using the methods described herein is illustrated. In this example, ***the tumor is a leukemia and may be detected from a nucleic acid sample taken from a leukocyte.*** A blood sample is taken from a patient during a routine visit to the clinic. The blood sample is centrifuged to separate the blood into plasma, buffy coat, and red blood cells. The plasma contains cfDNA and the buffy coat contains leukocytes (i.e., some of which may be leukemia cells). Nucleic acids are extracted from the plasma and the leukocytes, the nucleic acids are sequenced, and variants are detected in the samples. ***In this example, the sequence identified in the cfDNA is the “normal”, sample and the leukocytes are the “tumor sample.”*** The sequences are compared and variants that occur only in the leukemia are identified.

Ex. 1001 at 53:53-54-:3 (“Example 4: Detecting Variants of Leukemia by Tumor Versus Normal Analysis”). The only difference between this example – which is not inventive – and Ley is the use of cfDNA from the blood plasma as the normal sample, whereas Ley uses a skin biopsy, which is a trivial and non-inventive design choice that would have been well within the knowledge of a POSA. As explained above in Section V.E, all cells in the body, including both healthy and normal cells, release cfDNA. Because the cfDNA present in plasma constitutes a collection of DNA shed from all sources in the body, and because normal cells almost invariably outnumber tumor cells, the overwhelming majority of the cfDNA constitutes DNA from normal cells, allowing its use as a germline control for nontargeted WGS. (cfDNA from a tumor can be an informative biomarker when targeted sequencing is used to detect known tumor DNA variants). Example 4, like Ley, differs from the claim insofar as the tumor sample is blood (rather than tissue) in a leukemia, but confirms that the claimed application to solid tumors is merely an obvious variation of Ley’s method, as explained in further detail below.

2. Adaptations of Ley’s Approach to Other Cancers

170. Following Ley, many studies by many groups implemented the same general concept of (i) paired WGS of tumor and normal DNA samples in order to identify variants of interest and (ii) monitoring of such variants in the patient to track their cancer. I discuss a subset of such studies here.

171. For example, Meyerson reported that “[c]omplete sequencing of the genome of cancer tissue to high redundancy, using germline DNA sequence from the same patient as a comparison, has the power to discover the full range of genomic alterations – including nucleotide substitutions, structural rearrangements, and copy number alterations – using a single approach.” Ex. 1031 at 687. The first of seven references Meyerson cites for this proposition is Ley.

172. Similarly, Chapman (Ex. 1103) observed that “[a] powerful way to understand the molecular basis of cancer is to *sequence* either *the entire genome* or the protein-coding exome, *comparing tumour to normal from the same patient to identify the acquired somatic mutations*.” Ex. 1103 at 467. Importantly, Chapman establishes the broad applicability of paired tumor/normal WGS assays to identify tumor specific mutations (i.e., mosaic variants), and cites exemplary “[r]ecent reports,” including Ley, which “have described the sequencing of whole genomes from a single patient.” *Id.* With Ley clearly providing inspiration, Chapman “report[ed] the massively parallel sequencing of 38 tumour genomes and their comparison to matched normal DNAs.” *Id.* Specifically, Chapman analyzed 38 multiple myeloma patients using WGS of DNA “extracted from bone marrow aspirate (tumour) and blood (normal).” *Id.* at 471. Through this process, Chapman analyzed mosaic variants in “38 multiple myeloma patients, performing whole-genome sequencing (WGS) for 23 patients and whole-exome sequencing (WES;

assessing 164,687 exons) for 16 patients, with one patient analyzed by both approaches.” Ex. 1103 at 467 (Landscape of multiple myeloma mutations).

173. Another of the “recent reports” cited by Chapman is Lee 2010 (Note 5),⁸ which performed WGS on paired tumor and adjacent normal tissue of a lung cancer patient. Ex. 1104 at 473 (“Here we present the complete sequences of a primary lung tumour (60x coverage) and adjacent normal tissue (46x). Comparing the two genomes); *id.* at Table 1. Lee compared the two genomes to identify tumor-specific mosaic variants, including SNVs. *Id.* (“Comparing the two genomes, we identify a wide variety of somatic variations, including >50,000 high-confidence single nucleotide variants. We validated 530 somatic single nucleotide variants in this tumour, including one in the *KRAS* proto-oncogene and 391 others in coding regions, as well as 43 large-scale structural variations.”). The authors observed that “systematic comparison between a primary tumour and its matched normal genome gives a global view of the various forces shaping the complex mutation landscape of a solid tumor.” *Id.* at 476.

174. Yet another of the “recent reports” mentioned by Chapman is Ding 2010, which described the genomic analysis of four DNA samples from a breast

⁸ For its part, Lee also cites Ley as an example of “[r]ecent cancer sequencing efforts [that] have used next-generation sequencing technologies to provide a genome-wide view of mutations in leukemia, breast cancer and cancer cell lines.”

cancer patient implementing “[m]assively parallel DNA sequencing technologies” for its “unprecedented ability to screen entire genomes for genetic changes associated with tumour progression.” Ex. 1105 at 999. Ding compared its results to Ley’s⁹ given the similarity of the experimental protocols, finding that “this cancer genome [i.e., of the breast cancer patient], in comparison with the two AML (acute myeloid leukemia) cases published recently [citing Ley], revealed a 3-4-fold increase in high-confidence SNVs genome-wide, suggesting a much greater background mutation rate.” *Id.* at 1005.

175. For its “full genome sequencing,” Ding prepared DNA samples from the patient’s “peripheral blood, primary tumour, brain metastasis and an early passage xenograft.” *Id.* Ding performed WGS generating over 110 billion base pairs of sequence data from the “genomic DNA derived from blood, primary tumour, brain metastasis and xenograft samples, respectively, with corresponding haploid coverages of 38.8x, 29.0x, 32.0x and 23.8x.” *Id.* at 999-1000. Mosaic variants, including point mutations and small indels, were identified by comparing the sequencing data for the peripheral blood (normal, “N”) sample with the primary tumor (“T”), metastases (“M”) and xenograft (“X”) tumor. *Id.* at 1002 (Table 2). A POSA would have known that the peripheral blood comprises leukocytes. Ex. 1109

⁹ Ding and Ley have a number of overlapping co-authors, including Tim Ley and Elaine Mardis.

(Qiagen Protocol) at 29 (stating that “[a] 200 μ l sample of whole human bloods” contains a concentration of $\sim 5 \times 10^6$ leukocytes/ml). Moreover, a POSA would have found it obvious to extract DNA from buffy coat, which is “a leukocyte-enriched fraction of whole blood,” to obtain DNA for a sequencing experiment because “[p]reparing a buffy-coat fraction from whole blood is simple and yields approximately 5-10 times more DNA than an equivalent volume of whole blood.” *Id.*

176. These studies document the level of skill in the art, and establish that a POSA would have found it obvious to apply and extend the general method of Ley – identification of tumor-specific mutations based on comparison of paired tumor/germline WGS data followed by serial monitoring of such mutations over time – to a variety of cancers, including solid tumors. When working with cancers that present with solid tumors, a POSA would have known, and would have found it obvious, to use a tumor sample (i.e., a biopsy) as a source of tumor DNA, and a blood sample containing leukocytes as a source of germline DNA, which would have been an obvious and predictable variation of the methods initially disclosed by Ley.

3. Claim 1

177. The preamble of claim 1 recites “[a] method of analyzing a biological sample from a subject being screened for a cancer, wherein the biological sample is from a non-fetal source.”

178. Ley discloses a method of analyzing multiple biological samples using “massively parallel sequencing technology to sequence the genomic DNA of tumour and normal skin cells obtained from a patient” with leukemia. Ex. 1012 (Ley) at 66. Ley discloses analysis of samples from a subject being screened for cancer, given that the assay sequenced the patient’s tumor’s genome (at multiple times) and compared “its matched normal counterpart obtained from the same patient’s skin.” *Id.* Through this analysis, Ley “discovered ten genes with acquired mutations; two were previously described mutations that are thought to contribute to tumour progression, and eight were new mutations present in virtually all tumour cells at presentation and relapse. . . .” *Id.*

179. The method of Ley analyzes samples from a non-fetal source, as AML is a disease that effects adults. *Id.* at 66. Ley specifically discloses detailed analysis of biological samples from “a previously healthy woman in her mid-50s.” *Id.* at 67. Following Ley, studies confirmed that the same analysis applied to samples taken contemporaneously with the patients’ disease and treatment. *See e.g.*, Ding, Chan, Okosun-II.

180. The method disclosed by Ley meets each limitation of claim 1, as discussed below.

a) Step 1(a): [1] obtaining a first set of sequence reads and a second set of sequence reads by sequencing nucleic acid molecules derived from a first set of nucleic acid molecules and a second set of nucleic acid molecules, [2] wherein said first set of sequence reads corresponds to said first set of nucleic acid molecules and said second set of sequence reads corresponds to said second set of nucleic acid molecules, [3] wherein said first set of sequence reads and said second set of sequence reads are generated using whole genome sequencing, and [4] wherein (i) said first set of nucleic acid molecules is extracted from tissue sample of said subject, and (ii) said second set of nucleic acid molecules is extracted from leukocytes from a blood sample

181. Step 1(a) requires performing whole-genome sequencing (WGS) on nucleic acid molecules extracted from (i) a tissue sample from the subject; and (ii) leukocytes from a blood sample. Ley renders obvious step 1(a) by performing WGS on matched (i) a tissue sample in the form of a skin sample; and (ii) AML tumor DNA. As explained above, Ley's method is applicable to cancers which present as solid tumors (whereas Ley analyzed AML, a blood cancer), and in such case a POSA would have found it obvious to perform the WGS step on nucleic acid molecules extracted from (i) a tumor (tissue) sample from the subject; and (ii) leukocytes from a whole blood or buffy coat sample. Indeed, an example of such an approach was successfully implemented by Ding, which compared a breast cancer tumor sample with paired normal peripheral blood, which includes leukocytes, using a WGS assay. Ex. 1105. Ding thus confirms that a POSA would have had a reasonable expectation

of success in modifying Ley by altering the sample types to utilize the method in the screening of solid tumors, because the combination would be the simple application of known techniques to utilize similar, known sequencing and bioinformatics technology in the same way.

(1) Step 1(a)([1]): “obtaining a first set of sequence reads and a second set of sequence reads by sequencing nucleic acid molecules derived from a first set of nucleic acid molecules and a second set of nucleic acid molecules.”

182. Ley discloses the use of whole genome sequencing to sequence two sets of nucleic acid molecules from a subject to generate sequencing data, which includes sequencing reads. For example, Ley states that they “used massively parallel sequencing technology to sequence the genomic DNA of tumour and normal skin cells obtained from a patient with a typical presentation of French-American-British (FAB) subtype M1 acute myeloid leukemia (AML) with normal cytogenetics.” Ex. 1012 at 66. The sequencing reads obtained from the normal skin cells constitute a first set of sequencing reads, and the sequencing reads obtained from the tumor cells constitute a second set of sequencing reads. Ley’s Table 1 (excerpted below) depicts the number of sequencing reads obtained for each of the sample types. *Id.* at 67-68. Obtaining sequence reads from the samples necessarily involved sequencing nucleic acid molecules derived from the tumor and skin samples. *See* Section V.C.

Table 1 | Tumour and skin genome coverage from patient 933124

	Tumour	Skin
Libraries	4	3
Runs	98	34
Reads obtained	5,858,992,064	2,122,836,148
Reads passing quality filter	3,025,923,365	1,228,177,690
Bases passing quality filter	98,184,511,523	41,783,794,834
Reads aligned by Maq	2,729,957,053	1,080,576,680
Reads unaligned by Maq	295,966,312	138,276,594

183. Additionally, Ley’s “Methods Summary”: “Sequence end reads (average length for tumour genome, 32 bp, and for skin, 35 bp) were generated from Illumina/Solexa fragment libraries derived from the tumour or skin cells of patient 933124, using the Illumina Genome Analyser. The analysed reads were aligned to the human reference genome (NCBI Build 36) using Maq.” *Id.* at 71.

(2) Step 1(a)([2]): “wherein said first set of sequence reads corresponds to said first set of nucleic acid molecules and said second set of sequence reads corresponds to said second set of nucleic acid molecules”

184. As discussed above, Ley generates a first set of sequence reads as a result of sequencing nucleic acid molecules from the skin sample, and a second set of sequence reads as a result of sequencing nucleic acid molecules from the tumor sample. As a result, each set of sequence reads corresponds to the respective set of nucleic acid molecules from which they were generated using Ley’s whole-genome sequencing assay. *See* Ex. 1012 at 66-68, 71.

(3) Step 1(a)([3]): “wherein said first set of sequence reads and said second set of sequence reads are generated using whole genome sequencing,”

185. As noted above, Ley discloses that its sequencing reads are generated using whole genome sequencing. Specifically, Ley states that “[f]or the tumour genome, 32.7-fold ‘haploid’ coverage (98 billion bases) was obtained, and 13.9-fold coverage (41.8 billion bases) was obtained for the normal skin sample.” Ex. 1012 at 66; *see also id.* (“These results demonstrate the power of whole-genome sequencing to discover new cancer-associated mutations.”); *id.* at 71 (“In summary, we have successfully used a next-generation whole-genome sequencing approach to identify new candidate genes that may be relevant for AML pathogenesis.”).

(4) Step 1(a)([4]): “wherein (i) said first set of nucleic acid molecules is extracted from tissue sample of said subject, and (ii) said second set of nucleic acid molecules is extracted from leukocytes from a blood sample”

186. As discussed above, Ley discloses performing its whole-genome sequencing on (i) an AML tumor sample; and (ii) a skin sample as a paired source of germline DNA. As explained above, applying Ley’s methods to solid cancers, such as breast cancer tumors, would have been obvious to a POSA, and would have involved the simple application of known techniques. To do so, a POSA would have found it obvious to modify Ley’s method by (i) extracting the tumor nucleic acid molecules from a tissue sample affected by a cancer (e.g., a breast cancer tumor

biopsy); and (ii) extracting the paired normal nucleic acid molecules from leukocytes from a blood sample. This is confirmed by Ding's performance of such a method by 2011 using paired breast cancer tumor and peripheral blood. Ex. 1105; *see* Ex. 1109 at 29 (stating that "[a] 200 μ l sample of whole human bloods" contains a concentration of $\sim 5 \times 10^6$ leukocytes/ml). Moreover, a POSA would have found it obvious to extract DNA from buffy coat, which is "a leukocyte-enriched fraction of whole blood," to obtain DNA for a sequencing experiment because "[p]reparing a buffy-coat fraction from whole blood is simple and yields approximately 5-10 times more DNA than an equivalent volume of whole blood." *Id.*

b) "using said first set of sequence reads and said second set of sequence reads to identify mosaic variants specific to said tissue sample of said subject in said first set of sequence reads that is not present in said second set of sequence reads;"

187. Ley renders obvious step 1(b) by comparing the sequence reads from the tumor sample to the sequence reads from the germline sample to identify mosaic variants specific to the tumor that are not present in the normal, germline sample. As explained above, it would have been obvious to perform this analysis using a tissue sample as a tumor sample, and a blood sample comprising leukocytes as a germline sample, for example in the context of a solid tumor. *See* Ex. 1105 (Ding) (WGS of primary breast cancer tumor versus peripheral blood germline); Ex. 1031 (Meyerson) at 686 ("[t]he samples generally used for germline genome analysis [are]

peripheral blood mononuclear cells.”); *see also* Ex. 1008 (Chan) (paired HCC tumor/buffy coat normal WGS); Exs. 1101-1102 (Okosun and Okosun-II) (paired FL lymph node biopsy/peripheral blood WGS).

188. For example, Ley discloses that the sequencing revealed a large number of genetic variants in the tumor sample as compared to the normal, germline reference. In particular, a variety of “SNVs,” which are genetic variants, are described in Table 1, which identifies SNVs that were identified in the tumor sample, in the skin sample, and thus the corresponding number of SNVs that were detected in the tumor sample but not the skin sample (e.g., 3,811,115 SNVs detected with respect to hg18 in the tumor, 2,918,446 detected in the skin, and thus 892,669 mosaic variants present in the tumor but not the skin).¹⁰

¹⁰ Reference sequences (e.g., “hg18,” “Venter genome,” and “Watson genome”) are also listed.

Table 1 | Tumour and skin genome coverage from patient 933124

	Tumour	Skin
Libraries	4	3
Runs	98	34
Reads obtained	5,858,992,064	2,122,836,148
Reads passing quality filter	3,025,923,365	1,228,177,690
Bases passing quality filter	98,184,511,523	41,783,794,834
Reads aligned by Maq	2,729,957,053	1,080,576,680
Reads unaligned by Maq	295,966,312	138,276,594
SNVs detected with respect to hg18 (no Y)	3,811,115	2,918,446
SNVs (chr 1-22) detected with respect to hg18	3,681,968 (100.0%)	2,830,292 (100.0%)
SNVs also present in dbSNP	2,368,458 (64.3%)	2,161,695 (76.4%)
SNVs also present in Venter genome	1,499,010 (40.7%)	1,383,431 (48.9%)
SNVs also present in Watson genome	1,573,435 (42.7%)	1,456,822 (51.5%)
SNVs not in dbSNP/Venter/Watson	1,223,830 (33.2%)	591,131 (20.9%)
SNVs not in dbSNP/Venter/Watson/skin	925,200 (25.1%)	—
HQ SNPs	46,494 (100.0%)	46,572 (100.0%)
HQ SNPs where reference allele is detected	42,419 (91.2%)	38,454 (82.6%)
HQ SNPs where variant allele is detected	43,164 (92.9%)	39,220 (84.2%)
HQ SNPs where both alleles are detected	42,415 (91.2%)	38,454 (82.6%)

Assessments are shown of the haploid and diploid coverage of the tumour and skin genomes from AML patient 933124. Chr, chromosome; hg18, human genome version 18; HQ, high quality.

189. Ley subsequently filters SNV calls using its bioinformatics protocols to focus more specifically on a high-quality subset of tumor specific SNVs (mosaic variants) that were present in the tumor but not in the normal skin sample. Specifically, “[b]ecause the number of sequence variants initially detected by Maq was high, we developed improved filtering tools to effectively separate true variants from false positives.” Ex. 1012 at 67-68. Ley explains that “[w]e examined the Maq alignments by Decision Tree analysis to discover SNVs, as well as to identify copy number variants. Non-aligned reads were further analysed for indel discovery.” *Id.* at 71.

190. Ley compares the sequencing data from the tumor and normal samples to identify mosaic variants, and focused later analyses on ten mosaic variants in

genes that occur in the tumor sample but not the normal sample. (Ley refers to these mosaic variants as “acquired” or “somatic mutations”). *See, e.g.*, Ex. 1012 at 70.

191. Ley describes the identification of mosaic variants in the tumor sample in the section titled “Identification of somatic mutations in coding sequences.” Ex. 1012 at 68. Ley found that “[t]he patient had 3,813,205 sequence variants in her tumour genome,” which Ley filtered down to 2,647,695 using its bioinformatics workflow, and “2,584,418 (97.6%) were also detected in the skin genome (Fig. 2).” *Id.* Ley then performed further filtering on the 63,277 mosaic variants identified in the tumor sample but not in the skin sample, resulting in 8 validated SNVs (mosaic variants) for further study. *Id.* (Fig. 2).

192. Ley further described “Detecting insertions and deletions (indels)” to identify additional mosaic variants that were present in the tumor sample but not in the normal sample. Ex. 1012 at 69. Ley’s “efforts resulted in the identification of 726 putative small indels,” which were further reviewed and filtered to a set of 28 putative indels. *Id.* “Of these putative indels, 22 were validated but were found present in both tumour and skin..., two were false positive calls, two had no coverage, and two were previously validated somatic insertions in *NPM1* (4bp) and *FLT3* (30 bp).” *Id.* Thus, Ley identified 6 mosaic variants in the form of indels present in the tumor sample but not in the skin sample.

193. In further studies, Ley's focused on 10 mosaic variants comprising "[a] total of ten non-synonymous somatic mutations [that] were identified in this patient's tumour genome" but not in the normal tissue sample: the 8 validated SNVs and the two previously known indels in NPM1 and FLT3. Ex. 1012 at 70. The application of a similar bioinformatics workflow to sequencing data generated from WGS of paired solid tumor and normal blood/leukocytes – as in Ding – would have been an obvious application of known methods in a known way to achieve predictable results, namely the identification of mosaic variants present in a tumor sample that were not present in a matched germline sample.

c) Step 1(c): "subsequent to (b), identifying said mosaic variants specific to said tissue sample of said subject in a first additional sample of said subject, wherein said first additional sample is obtained from a source that is different than said tissue sample and at a time point that is different than said tissue sample;"

194. Ley renders obvious identifying tumor-specific mosaic variants from an additional tumor sample that was obtained from a different source at a different time as compared to the initial tumor sample. Specifically, Ley discloses identifying the tumor-specific mosaic variants identified in step 1(b) in a relapse tumor sample.

195. Ley discloses analyzing an additional biological sample of the subject, *e.g.*, a sample from a "relapse tumor," to generate a sequencing library, a process in which Ley specifically focuses on identifying the aforementioned validated mosaic

variants (such as the eight validated SNPs). For example, Ley uses PCR probes corresponding to the genetic variants identified in the patient tumor plus two additional PCR probes corresponding to exons in cancer-associated genes known to have common genetic variants to selectively amplify sequences from both the original tumor and skin samples as well as one additional sample obtained from the patient's relapsed disease, and to thereby generate amplicons comprising a sequencing library. "[W]e amplified each mutation containing locus from non-amplified genomic DNA derived from the *de novo* and relapse tumour samples, and from the skin biopsy obtained at presentation. The resulting amplicons were sequenced using the Roche/454 FLX platform, and the frequency of reads containing the reference and variant alleles were defined (Fig. 3 and Table 3)." Ex. 1012 at 69.

196. Figure 3 summarizes this data, and shows the frequency of mosaic variants ("ten somatic mutations") in each of the primary tumor, relapse tumor, and skin samples.¹¹ This shows that Ley identified the presence of the mosaic variants (green bars) specific to the tumor sample in the subsequently obtained relapse tumor sample.

¹¹ Ley notes that the skin sample shows the presence of certain of the MVs because "the normal skin sample was contaminated with leukaemic cells containing the somatic mutations." Ex. 1012 at 69 (Fig. 3 caption).

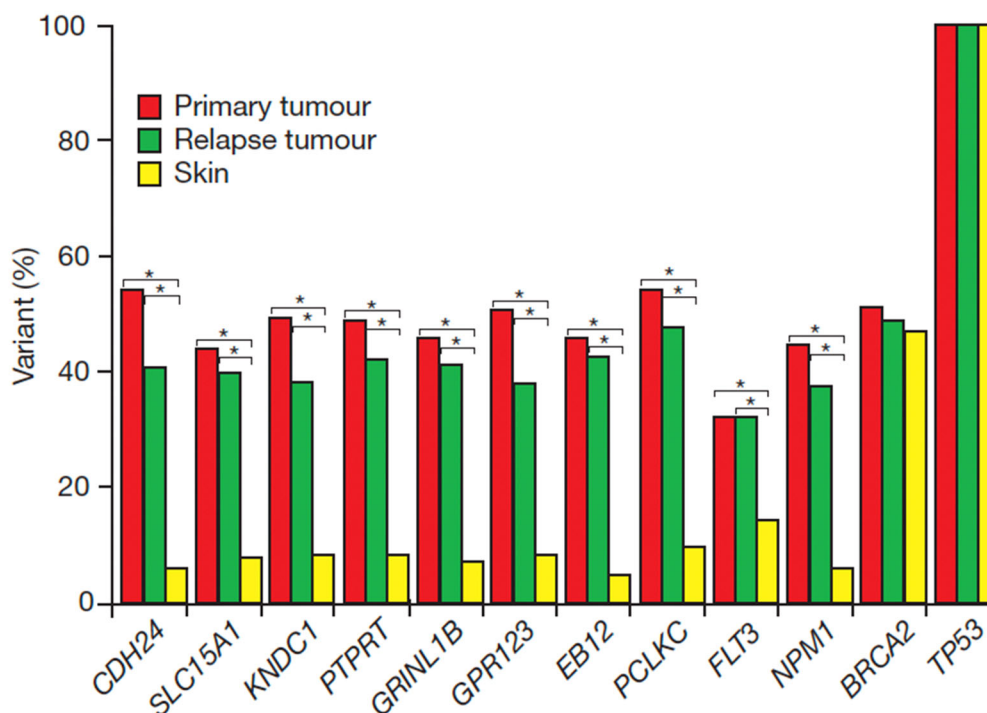


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that of the skin sample for all mutations ($P < 0.000001$ for all mutations, Fisher's exact test, denoted by a single asterisk in all cases). Note that the normal skin sample was contaminated with leukaemic cells containing the somatic mutations. The patient's white blood cell count was 105,000 (85% blasts) when the skin punch biopsy was obtained.

d) Step 1(d): providing a report based on an identification of said mosaic variants specific to said tissue sample of said subject in said first additional sample, wherein said report comprises predicting, diagnosing, and/or prognosing a status or outcome of a cancer in said subject.

197. Ley further discloses generating a report based on the identification of the frequencies of the mosaic variants occurring in the primary and relapse tumor samples. Specifically, Ley's Figure 3 reports sequence information regarding the

relative abundance of particular somatic and germline mutations that were observed in samples taken at (i) primary onset, and (ii) relapse. Ex. 1012 (Ley) at 69 (Fig. 3) (reproduced below). These abundances are compared to the frequency of the same polymorphic alleles present in the skin sample, which represents a control. This report comprises predicting, diagnosing, and/or prognosing the status of the cancer at least insofar as it prognoses a status – the variant fraction of the identified mosaic variants – of cancer cells in the patient. For example, a POSA would have understood that if the second sample did not show the presence of the tumor-associated mosaic variants (and thus appeared similar to the germline sample), that such a result might be consistent with remission or an improved condition.

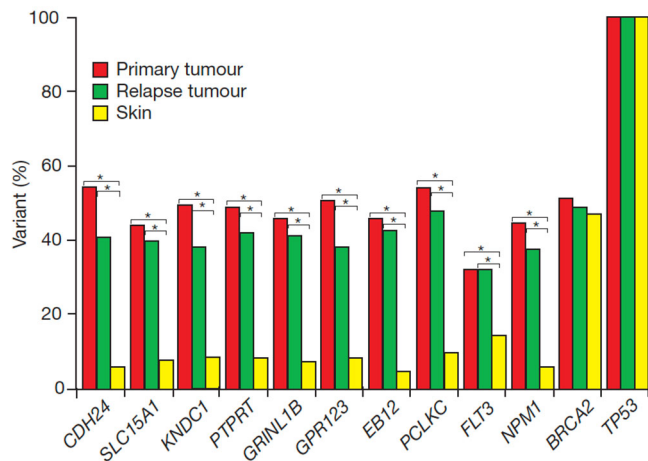


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that of the skin sample for all mutations ($P < 0.000001$ for all mutations, Fisher's exact test, denoted by a single asterisk in all cases). Note that the normal skin sample was contaminated with leukaemic cells containing the somatic mutations. The patient's white blood cell count was 105,000 (85% blasts) when the skin punch biopsy was obtained.

198. Thus, Ley in view of knowledge of a POSA renders claim 1 obvious.

4. Claim 2

199. Claim 2 recites, “[t]he method of claim 1, further comprising identifying mosaic variants specific to said tissue sample of said subject in a second additional sample of said subject, wherein said second additional sample comprises cell-free DNA, which second additional sample is obtained at a time point that is different than said tissue sample and said first additional sample.”

200. As explained with respect to claim 1, including step 1(c), above, Ley in view of the knowledge of a POSA renders obvious analyzing various biological samples in cancer patients by (i) performing whole genome sequencing on paired tumor tissue and a source of germline DNA, such as a blood sample containing leukocytes; (ii) identifying tumor-specific mosaic variants (such as SNVs and indels present in tumor DNA but not in germline DNA) in the first-obtained tumor sample; and (iii) identifying those tumor-specific mosaic variants in additional patient samples, such as later-arising tumor samples (e.g., relapse or metastasized tumor samples), including to determine the patient’s ongoing disease burden, disease status, and/or response to treatment.

201. A POSA would have known that circulating cell-free DNA, such as that present in the blood, and specifically in the blood plasma, was useful as a biomarker for multiple types of cancer. *See* Section V.E. When paired with knowledge of

tumor-associated SNVs and/or indels (or other mutations identified through an initial whole-genome sequencing step), a POSA would have been motivated to detect such tumor-specific mosaic variants in additional blood samples taken over time in order to serially monitor the status of the patient's cancer. Exemplary studies that successfully demonstrated this approach included Chan, Dawson, Leary, and McBride, each discussed in turn.

202. **Chan:** As discussed above in Grounds 1 and 2, Chan disclosed serially monitoring mosaic variants in cell-free DNA in plasma circulation after identifying them in paired tumor/normal whole genome sequencing. Chan further renders obvious the use of capture probes, such as solution-based hybridization capture probes for the targeted detection of mosaic variants of interest. Ex. 1008.

203. **Dawson (Ex. 1070):** Dawson discloses a study evaluating “circulating cell-free DNA carrying tumor-specific alterations (circulating tumor DNA)” as a “circulating biomarker[] in breast cancer.” Ex. 1070 at 1199 (Background). One aspect of Dawson's investigation was to use “whole-genome sequencing to identify somatic genomic alterations and design[] personalized assays to quantify circulating tumor DNA in serially collected plasma specimens.” *Id.* Dawson reported that “[c]irculating tumor DNA was successfully detected in 29 of the 30 women (97%) in whom somatic genomic alterations,” i.e., mosaic variants identified in the tumor but not in the germline, “were identified,” concluding that “circulating tumor DNA

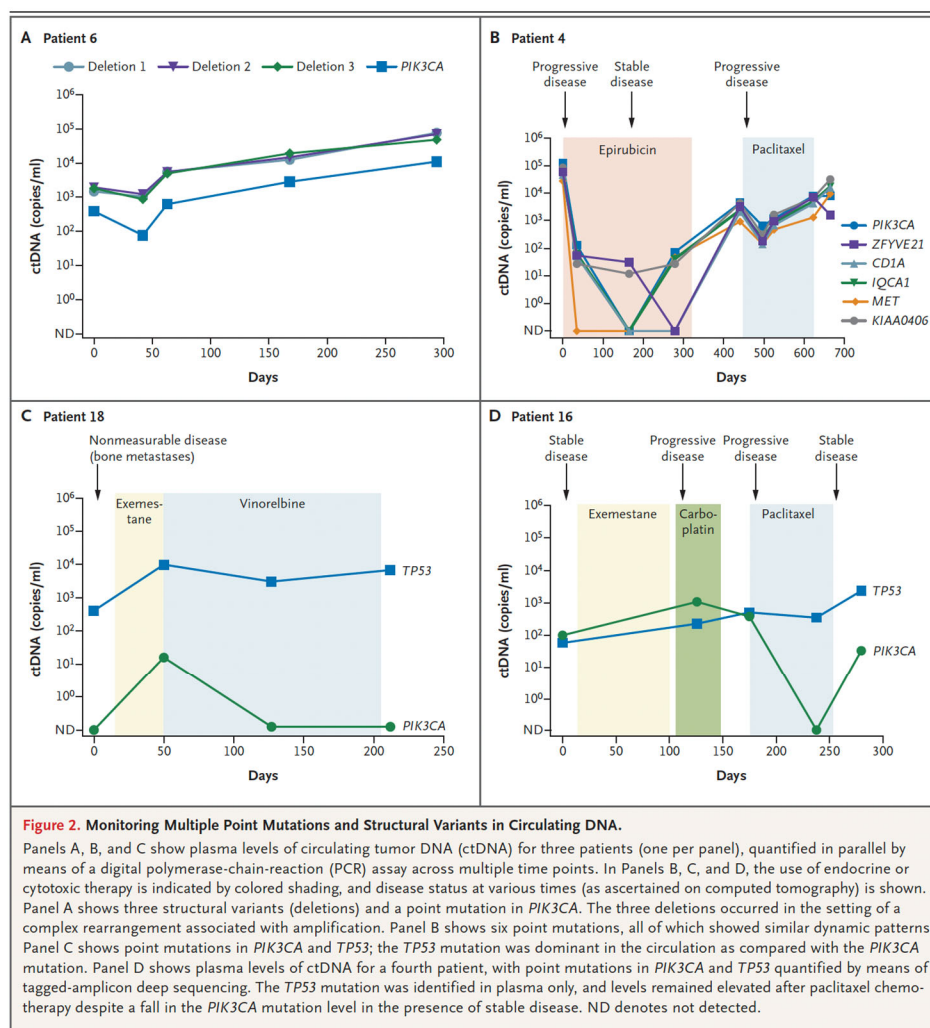
is an informative, inherently specific, and highly sensitive biomarker of metastatic breast cancer.” *Id.*

204. The patients being screened for cancer in Dawson “were women with metastatic breast cancer currently undergoing active treatment,” of which “30 had genomic alterations suitable for monitoring.” *Id.* at 1200. Such monitoring involved the analysis of “[s]erial blood samples (30 ml each) . . . at intervals of 3 or more weeks.” *Id.* For the “identification of somatic genomic alterations,” Dawson explains that “[s]equencing was performed on DNA from breast-cancer specimens and matched normal tissue specimens, with the use of one or both of two methods,” one of which was “paired-end whole-genome sequencing.” Dawson’s supplementary appendix, consistent with the discussion above, confirms that leukocytes are an appropriate source of matched normal DNA. Ex. 1117 at 6 (“In parallel to the processing of tumor tissues, matched normal DNA was obtained from all study participants to enable confirmation of somatic versus germline genetic events. This was derived from normal peripheral blood leucocytes (NPBL).”). Using whole-genome sequencing, Dawson “identified somatic structural variants in 8 patients,” and “[i]n 3 patients, both mutations and structural variants were identified, enabling us to compare and contrast the use of point mutations and structural variants for serial monitoring of circulating tumor DNA.” *Id.* at 1201-1202, Fig. 1. Moreover, Dawson used whole genome sequencing in one patient “to identify multiple somatic

mutations, enabling us to monitor multiple mutations in parallel in circulating tumor DNA.” *Id.* Thus, Dawson’s whole-genome sequencing was used to identify mosaic variants, including point mutations (SNPs) and structural variants.

205. After identifying mosaic mutations in the tumor sample, Dawson then analyzed serial blood samples to identify such mosaic variants, in the plasma using two different techniques to interrogate the cell free circulating tumor DNA: digital PCR and TAm-Seq. *Id.* at 1202-1203. Figure 2 in Dawson constitutes a report that shows the plasma levels of the tumor-specific mosaic variants in four different patients, including reports on their disease status and response to therapeutic intervention. *Id.* at 1203.

CIRCULATING TUMOR DNA IN METASTATIC BREAST CANCER



206. **Leary: (Ex. 1002) and McBride (Ex. 1015):** Leary (Ex. 1002) and McBride (Ex. 1015) demonstrated the use of the same concept, by using whole-genome sequencing to identify structural variants, and to subsequently use those identified markers as targets for serial monitoring of tumor DNA in the blood plasma of patients with solid tumors. *See* Ex. 1002 at 5 (explaining that the identified mosaic variants (structural variants, which constitute indels) “can be used to detect tumor

DNA in the presence of large quantities of normal DNA and in patient plasma.”); Ex. 1015 at 8 (“Tumor-specific genetic variants,” i.e., mosaic variants, “unequivocally identify cancer cells.”); *id.* at 9 (explaining that Leary, like McBride “shows that tumor-specific rearrangements,” which are mosaic variants, “can be sensitively and specifically detected in plasma, with serial studies in one patient showing that levels correlated with decreasing disease burden during treatment. Our study adds to these findings by extending the proof-of-principle into two other tumor types, by showing that assays are positive at disease relapse as well as initial diagnosis and that impending relapse can potentially be identified.”).

207. When applying the collective teachings of Ley and the many successor studies that applied its core approach of determining tumor-specific variants (i.e., mosaic variants) via matched-normal WGS, followed by targeted serial monitoring, a POSA therefore would have been motivated to identify mosaic variants that were identified by whole genome sequencing of a tumor sample in serial plasma samples taken over time. There are multiple reasons for this, including to monitor disease burden, analyze the patient’s response to therapy or other medical intervention, and to predict impending relapse.¹² A POSA would have reasonably expected success in performing such monitoring in view of the multiple successful reports that did so,

¹² *See, e.g.*, Exs. 1002, 1008, 1015, 1070.

using untargeted shotgun approaches (Chan), targeted approaches using solution phase hybridization capture probes (Chan and cited references), and PCR-based capture techniques (Dawson, Leary, and McBride).

208. Thus, Ley in view of the knowledge of a POSA renders claim 2 obvious.

5. Claim 3

209. Claim 3 recites, “[t]he method of claim 2, wherein said first and/or second additional sample comprises a blood sample, a saliva sample, a buccal swab, a urine sample, a semen sample, or a cerebrospinal fluid sample of said subject.”

210. As explained with respect to claim 2, above, the use of serial plasma samples as a source of cell-free tumor DNA comprising mosaic variants is rendered obvious by Ley in view of the knowledge of a POSA.

211. Thus, Ley in view of the knowledge of a POSA renders claim 3 obvious.

6. Claim 4

212. Claim 4 recites, “[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are selected from the group consisting of: copy number variation, small insertions and deletions (inDels), single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), and any combination thereof.”

213. As explained above with respect to claims 1 and 2, Ley discloses identifying mosaic variants to the tumor tissue sample in the form of tumor specific single nucleotide variants, which are also known as single nucleotide polymorphisms. Ley also discloses the analysis of at least inDels.

214. Thus, Ley in view of the knowledge a POSA renders claim 4 obvious.

7. Claims 5 and 6

215. Claim 5 recites, “[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are causal or predisposing for a cancer.”

216. Claim 6 recites, “[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are non-causal for a cancer.”

217. The ‘968 patent does not disclose any information regarding the identification of a particular mosaic variant as “causal” or “non-causal,” beyond the tautology that a mosaic variant may be either causal or non-causal. *See, e.g.*, Ex. 1001 at 12:15-17 (“The specific mosaic variants can be used as biomarkers . . . even if they are not causal for medical condition.”); *id.* at 13:38-40 (“In some cases, the tumor variants are causal variants (i.e., tumor-causing).”). The specification does not provide any criteria for when a variant should be considered “causal” or not.

218. At a minimum, the mosaic variants identified in Ley and other references reflecting the knowledge of a POSA necessarily constitute either causal or non-causal mutations. Ley further explains that a POSA would have found it

obvious to determine whether detected mosaic variants are “passenger” mutations or alternatively mutations “truly important for pathogenesis (that is, disease ‘drivers’ versus passengers).” Ex. 1012 at 70. Moreover, as noted above, Chan explained that a POSA would have wanted to identify both passenger mutations and driver mutations. The passenger mutations “provide[] a numerical advantage that has the potential for more sensitive, precise, and representative detection of tumor DNA in plasma,” aiding tumor burden assessments, and also would have aided “selective analyses of driver mutations may allow more readily the identification of medically actionable targets.” Ex. 1008 at 222.

219. Accordingly, Ley in view of the knowledge of a POSA renders obvious detecting both causal and non-causal mosaic variants, and thus claims 5 and 6.

8. Claim 8

220. Claim 8 recites, “[t]he method of claim 1, further comprising providing a recommendation for a therapeutic intervention based on an identification of said mosaic variants specific to said tissue sample of said subject.”

221. The teachings of Ley in view of the knowledge of a POSA render obvious providing a recommendation for a therapeutic intervention based on an identification of said mosaic variants specific to said tissue sample of said subject. For example, in Dawson’s Figure 2, Dawson monitored the mosaic variants specific

to the tissue sample over time in serial plasma samples, noting the relationship between disease status and the administration of different therapies.

222. Accordingly, Ley in view of the knowledge of a POSA renders claim 8 obvious.

9. Claim 9

223. Claim 9 recites, “[t]he method of claim 1, wherein said tissue sample comprises a metastatic tissue sample, a mixed tissue sample, a biopsy, or a core needle biopsy.”

224. As explained above, the Ley, as well as subsequent references like Ding, use biopsies as tissue samples.

225. Thus, Ley in view of knowledge of a POSA renders Claim 9 obvious.

10. Claim 11

226. Claim 11 recites, “[t]he method of claim 1, wherein said mosaic variants specific to said subject are selected from the group consisting of: copy number variation, small insertions and deletions (inDels), single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), and any combination thereof.”

227. As explained above, the mosaic variants identified in Ley comprise at least single nucleotide variants or polymorphisms.

228. Thus, Ley in view of knowledge of a POSA renders claim 11 obvious.

11. Claim 12

229. Claim 12 recites, “[t]he method of claim 1, wherein said predicting, diagnosing, and/or prognosing of (d) comprises one or more of the following: (i) diagnosing a cancer, (ii) identifying a cancer, (iii) determining a stage of a cancer, (iv) assessing risk of a cancer, (v) assessing risk of cancer recurrence, (vi) assessing reproductive risk, (vii) assessing efficacy of a drug, (viii) assessing risk of an adverse drug reaction, (ix) predicting optimal drug dosage, (x) predicting drug resistance, or any combination thereof.”

230. The teachings of Ley in view of the knowledge of a POSA render obvious providing a recommendation for a therapeutic intervention based on an identification of said mosaic variants specific to said tissue sample of said subject. For example, Dawson monitored the mosaic variants specific to the tissue sample over time in serial plasma samples, noting the relationship between disease status and the administration of different therapies, thus obviating at least claim options (i); (ii); (iii); (vii) and combinations thereof. Ex. 1070 (Dawson Figure 2).

231. Accordingly, Ley in view of the knowledge of a POSA renders claim 12 obvious.

12. Claims 7, 10, and 13

232. Claim 7 recites, “[t]he method of claim 2, further comprising enriching said extracted nucleic acid molecules of said tissue sample, blood sample, first additional sample, or second additional sample.”

233. Claim 10 recites, “[t]he method of claim 1, further comprising enriching said extracted nucleic acid molecules of said tissue sample, blood sample, or first additional sample.”

234. Claim 13 recites, “[t]he method of claim 1, wherein said mosaic variants specific to said tissue sample of said subject are identified using capture probes in said first additional sample of step (c).”

235. As explained above with respect to claim 2, a POSA would have found it obvious to identify tumor-specific mosaic variants in serially obtained blood samples using techniques that identify said mosaic variants using capture probes, including PCR-based and solution phase capture probes. Such methods further constitute enriching the extracted nucleic acid molecules of the serially obtained additional samples.

236. Thus, Ley in view of knowledge of a POSA renders claims 7, 10, and 13 obvious.

D. Ground 4: Claims 1, 8-9, and 11-13 are Anticipated by Okosun-II

237. Claims 1, 8-9, and 11-13 are anticipated by Okosun-II. Ex. 1102. I note that the key findings of this paper were previously reported by the same group in a *Blood* abstract published November 16, 2012. Ex. 1101 (Okosun-I). Though my anticipation analysis below is based solely on Okosun-II, I have included parallel citations to note where Okosun-I disclosed the same substance in 2012.

1. Claim 1

238. The preamble of claim 1 recites “[a] method of analyzing a biological sample from a subject being screened for cancer, wherein the biological sample is from a non-fetal source.”

239. Specifically, Okosun-II performs serial WGS assays of follicular lymphoma (“FL”) biopsies and paired germline DNA:

An in-depth characterization and chronicling of the underlying genetic events from follicular lymphoma to transformed follicular lymphoma will serve as a guide in the development of targeted therapies. To address this need, we conducted whole-genome or whole-exome sequencing of sequential follicular lymphoma-transformed follicular lymphoma (FL-tFL) pairs and matched germline samples from 10 cases (Fig. 1a and Online Methods), with deep targeted sequencing of 28 genes in an extension cohort (Supplementary Fig. 1 and Supplementary Table 1).

Ex. 1102 at 176. Thus, Okosun-II discloses a method of screening subjects for cancer by analyzing biological samples from a non-fetal source.¹³

240. Moreover, as explained below, Okosun-II discloses the claimed method steps.

a) Step 1(a): [1] obtaining a first set of sequence reads and a second set of sequence reads by sequencing nucleic acid molecules derived from a first set of nucleic acid molecules and a second set of nucleic acid molecules, [2] wherein said first set of sequence reads corresponds to said first set of nucleic acid molecules and said second set of sequence reads corresponds to said second set of nucleic acid molecules, [3] wherein said first set of sequence reads and said second set of sequence reads are generated using whole genome sequencing, and [4] wherein (i) said first set of nucleic acid molecules is extracted from tissue sample of said subject, and (ii) said second set of nucleic acid molecules is extracted from leukocytes from a blood sample

241. Step 1(a) requires performing whole-genome sequencing (WGS) on nucleic acid molecules extracted from (i) a tissue sample from the subject; and (ii) leukocytes from a blood sample. Okosun-II discloses step 1(a) by performing WGS on matched (i) tumor tissue from FL lymph node biopsies; and (ii) paired germline sample. *See, e.g.*, Ex. 1102 at 176; *id.* at 182 (“**Case samples.** Ten individuals with

¹³ Ex 1101 (“We performed whole genome sequencing (WGS) using 100bp paired-end reads on tumor DNA of lymph node (LN) biopsies from 6 patients who were selected on the basis of availability of a FL, subsequent tFL biopsy and paired germline sample (8 FL, 6 tFL, 6 germline).”).

follicular lymphoma were selected on the basis of the availability of sequential tumor lymph node biopsies . . . and matched germline samples consisting of remission bone marrows or peripheral blood specimens.”).¹⁴

(1) Step 1(a)([1]): “obtaining a first set of sequence reads and a second set of sequence reads by sequencing nucleic acid molecules derived from a first set of nucleic acid molecules and a second set of nucleic acid molecules.”

242. Okosun-II discloses Step 1(a)([1]) by generating the claimed sequence reads: “[W]e conducted whole-genome or whole-exome sequencing of sequential follicular lymphoma-transformed follicular lymphoma (FL-tFL) pairs and matched germline samples from 10 cases.” Ex. 1102 at 176. Thus, Okosun-II obtained sequence reads by sequencing a first set of nucleic acid molecules derived from the FL tumor, and a second set of nucleic acid molecules derived from the paired germline sample. “For the ten index cases (six whole-genome sequenced and four whole-exome sequenced), the sequencing yielded mean genomic and exonic coverage of 37x and 110x, respectively, with 96% of the targeted bases covered by >10-fold.” *Id.*

¹⁴ Each subpart is disclosed in Okosun-I’s whole-genome sequencing of FL biopsies, tFL biopsies, and paired germline samples to 37x coverage of the mapped individual reads. Ex. 1101 (“We performed whole genome sequencing (WGS) using 100bp paired-end reads on tumor DNA of lymph node (LN) biopsies from 6 patients who were selected on the basis of availability of a FL, subsequent tFL biopsy and paired germline sample (8 FL, 6 tFL, 6 germline).”)

(2) Step 1(a)([2]): “wherein said first set of sequence reads corresponds to said first set of nucleic acid molecules and said second set of sequence reads corresponds to said second set of nucleic acid molecules”

243. The sequence reads obtained in Step 1(a)([1]) discussed above were necessarily obtained by sequencing DNA molecules, *see* Section Section V.C, and the sequence reads necessarily correspond to the sample from which they were obtained. In other words, the first set of sequence reads corresponds to the nucleic acid sequences of the tumor DNA molecules and the second set of sequence reads corresponds to the nucleic acid sequences of the paired germline DNA molecules.

(3) Step 1(a)([3]): “wherein said first set of sequence reads and said second set of sequence reads are generated using whole genome sequencing,”

244. As noted above, Okosun-II discloses that the first and second sets of sequence reads are generated using whole genome sequencing: “[W]e conducted whole-genome or whole-exome sequencing of sequential follicular lymphoma-transformed follicular lymphoma (FL-tFL) pairs and matched germline samples from 10 cases.” Ex. 1102 at 176; *see also id.* (“[Whole-genome] sequencing yielded mean genomic [] coverage of 37x [] with 96% of the targeted bases covered by >10 fold.”); *id.* at 182 (“**Whole-genome sequencing.** Genomic DNA (5 µg) from matched tumor and germline samples was fragmented to generate libraries with an

average insert size of 300 bp and sequenced using the HiSeq 2000 platforms with paired-end reads of 100 bp, according to the manufacturer's protocols (Illumina).").

(4) Step 1(a)([4]): "wherein (i) said first set of nucleic acid molecules is extracted from tissue sample of said subject, and (ii) said second set of nucleic acid molecules is extracted from leukocytes from a blood sample"

245. Okosun-II discloses that the first set of nucleic acid molecules is extracted from a lymph node biopsy. Ex. 1102 at 176-177, Fig. 1; *see also id.* at 182. Okosun-II further discloses that the germline DNA sample may be obtained from "peripheral blood specimens." *Id.* at 182 ("Ten individuals with follicular lymphoma were selected on the basis of the availability of sequential tumor lymph node biopsies (a sample with follicular lymphoma and a subsequent sample with histological transformation to DLBCL (transformed follicular lymphoma)) and matched germline samples consisting of remission bone marrows or peripheral blood specimens."). As explained above, peripheral blood specimens contain leukocytes, and thus satisfy the claim.¹⁵

¹⁵ While Okosun-I was silent on the source of its matched germline DNA, a POSA would have understood that normal blood, including fractions like buffy coat leukocytes, were one of the few sources of matched germline DNA that a POSA would at once envisage using in light of the disclosure of paired tumor/normal WGS.

b) Step 1(b): “using said first set of sequence reads and said second set of sequence reads to identify mosaic variants specific to said tissue sample of said subject in said first set of sequence reads that is not present in said second set of sequence reads;”

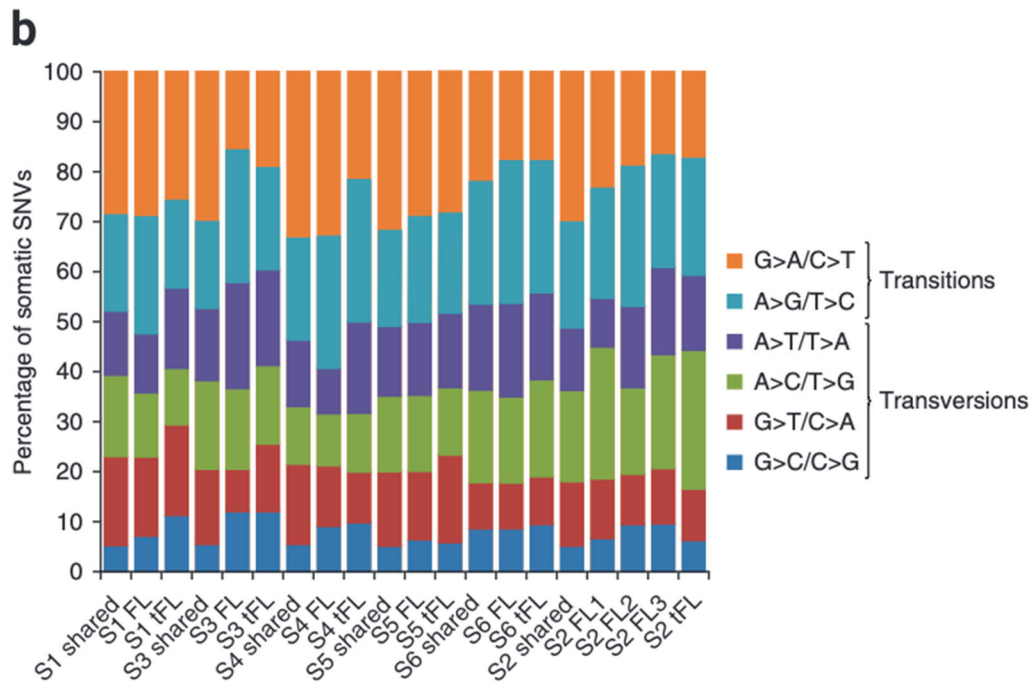
246. Okosun-II discloses identifying variants that were present in the FL genome (i.e., the genome of the FL lymph node biopsy) that were mutated as compared to the germline, thereby identifying mosaic variants specific to the FL tissue biopsy. For example, when describing its analysis of the generated WGS data (“**Whole-genome sequencing read mapping and somatic mutation detection**”), Okosun-II discloses: “We implemented the Strelka algorithm to identify somatic SNVs and indels for matched tumor-germline samples, selecting this algorithm for its sensitivity of mutation detection with varying tumor purities. After candidate indel detection and realignment, we called somatic variants and performed filtering.” Ex. 1102 at 182. The Strelka algorithm is well known in the art as “a method for somatic SNV and small indel detection from sequencing data of matched tumor—normal samples.” Ex. 1118 at 1811. Put simply, the Strelka algorithm compares the WGS data for the tumor sample with the germline sample to detect variations (SNPs and indels) that are present in one sample but not the other.¹⁶

¹⁶ Okosun-I discloses identifying mosaic variants from the whole-genome sequencing data it generated. Ex. 1101 (“Each tumor had an average of 72 somatic mutations (range: 41–143) with higher frequencies in the tFL genomes (mean: FL 64.8; tFL 83.6), indicating ongoing genetic evolution with disease progression.”).

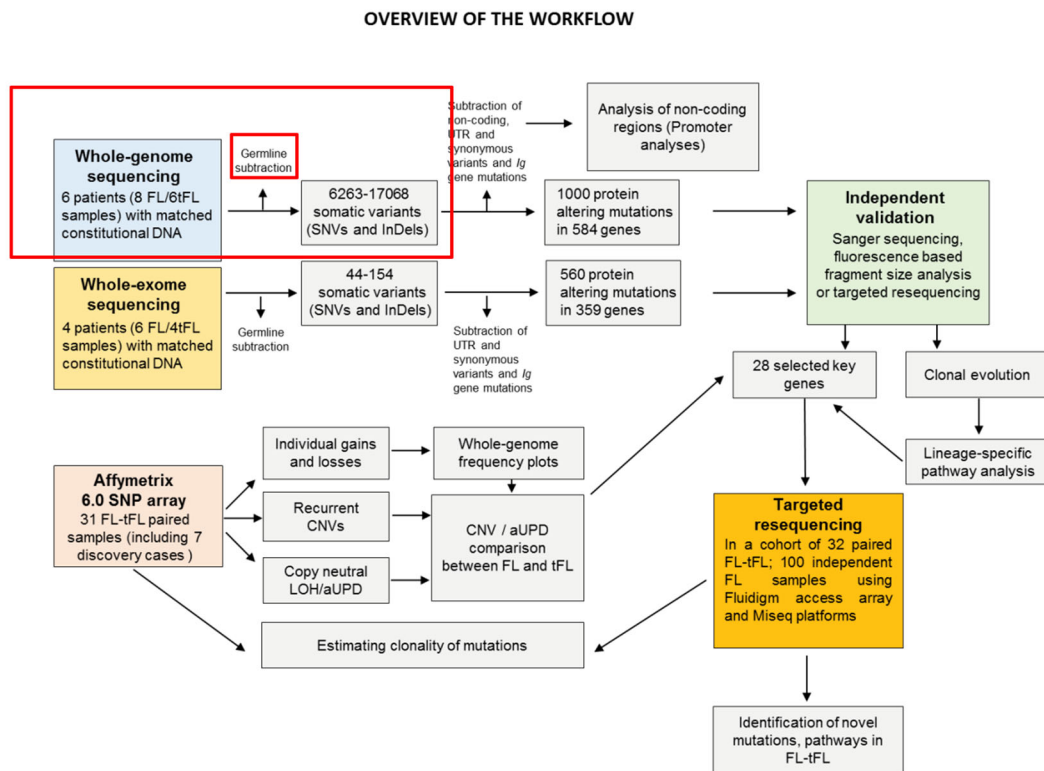
247. Okosun-II also describes additional filtration of mosaic variants for further analysis, noting that “[o]nly somatic calls originating from homozygous reference alleles in normal samples were considered.” *Id.* Subsequently, “[a]ll somatic SNVs and short indels retained after filtering were annotated,” and “allele frequencies for individual variants in both germline and tumour samples were determined by read count.” *Id.*

248. Figure 1b shows the distribution of base substitution patterns in the identified mosaic variants. “Somatic SNVs for each case were sorted into three categories: shared (variants present in both follicular and transformed follicular lymphoma samples), follicular lymphoma specific and transformed follicular lymphoma specific.” *Id.* at 177 (Figure 1b). Thus, mutations that are indicated as “shared” were present in both the FL and tFL tumor samples but not in the germline samples, while mutations that are indicated as “FL” were present in the FL sample but not in the germline (or subsequent tFL) samples.¹⁷

¹⁷ Patients S1-S6 were the patients whose tumors were subjected to WGS. Ex. 1113 (Supplementary Information) at 20 (Supplementary Table 1).



249. Supplementary Figure 1 (Ex. 1113) further evidences Okosun-II’s identification of mosaic variants by comparing the FL tumor and germline sequencing. Specifically, the step of “germline subtraction” from the whole-genome sequencing in matched tumor and normal (here referred to as “constitutional DNA”) refers to the identification of SNVs and indels – which constitute mosaic variants under Patent Owner’s understanding of that term – that were present in tumor sample but not in germline DNA.

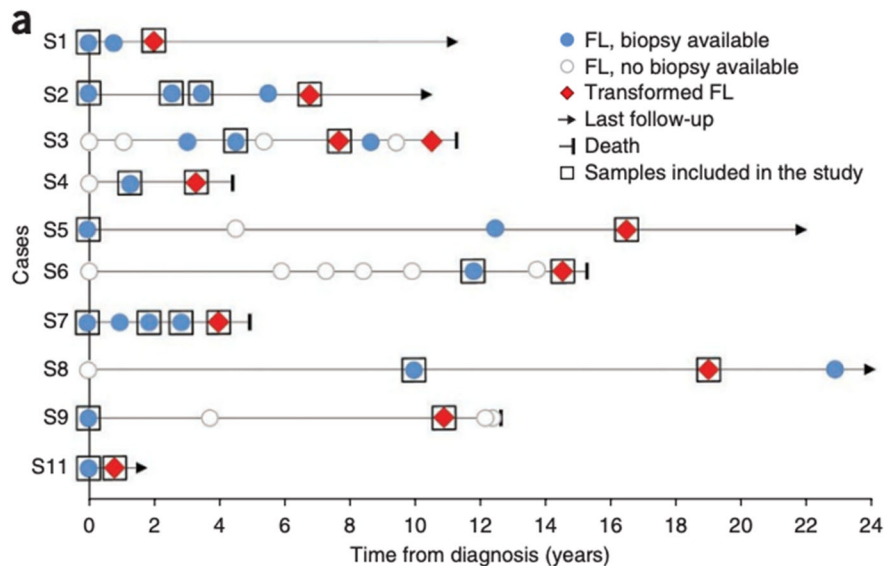


Supplementary Figure 1. Summary of study workflow. The schema shows the main analytical approaches used in this study: whole genome or exome sequencing; copy number profiling using Affymetrix SNP 6.0 arrays and targeted resequencing of candidate genes for prevalence screening.

c) Step 1(c): “subsequent to (b), identifying said mosaic variants specific to said tissue sample of said subject in a first additional sample of said subject, wherein said first additional sample is obtained from a source that is different than said tissue sample and at a time point that is different than said tissue sample;”

250. Okosun-II identifies mosaic variants specific to the FL tissue sample in a first additional sample that is obtained from a different source at a different time point. Specifically, Okosun-II performs serial biopsies on FL patients for additional FL samples, which include instances of continued FL (blue circle), as well as tFL

(red diamond). Samples that were analyzed in the study are boxed. Ex. 1102 at 177 (Fig. 1a).¹⁸



251. In patients S1-S6, Okosun-II discloses analyzing a transformed FL sample that was obtained after one or more FL samples was taken. *Id.* Okosun-II identifies the previously identified tumor-specific mosaic variants in the tFL samples, including in Figure 1b, where it identifies “shared” variants that were “variants present in both follicular lymphoma and transformed follicular lymphoma samples.” *Id.* at 177 (Fig. 1b).

¹⁸ Okosun-I discloses identifying mosaic variants that were present in both the FL and subsequent tFL samples. Ex. 1101 (“In all cases, we show an overlap of somatic variants (based on comparisons between non-synonymous, synonymous and UTR variants) between the paired FL/tFL biopsies, although this clonal relationship varied across the patients.”).

252. Through this analysis, Okosun-II “identified recurrent mutations,” i.e., tumor specific mosaic variants, “in linker histone, JAK-STAT signaling, NF-kB signaling and B cell developmental genes.” Ex. 1102 at 176. These results “suggest[ed] that personalizing therapies to target key genetic alterations in the CPC,” or “common progenitor clone,” “represents an attractive therapeutic strategy.” *Id.* In this analysis, Okosun-II describes the creation of developmental maps to track the evolution of mosaic variants in the follicular lymphoma samples to their transformation:

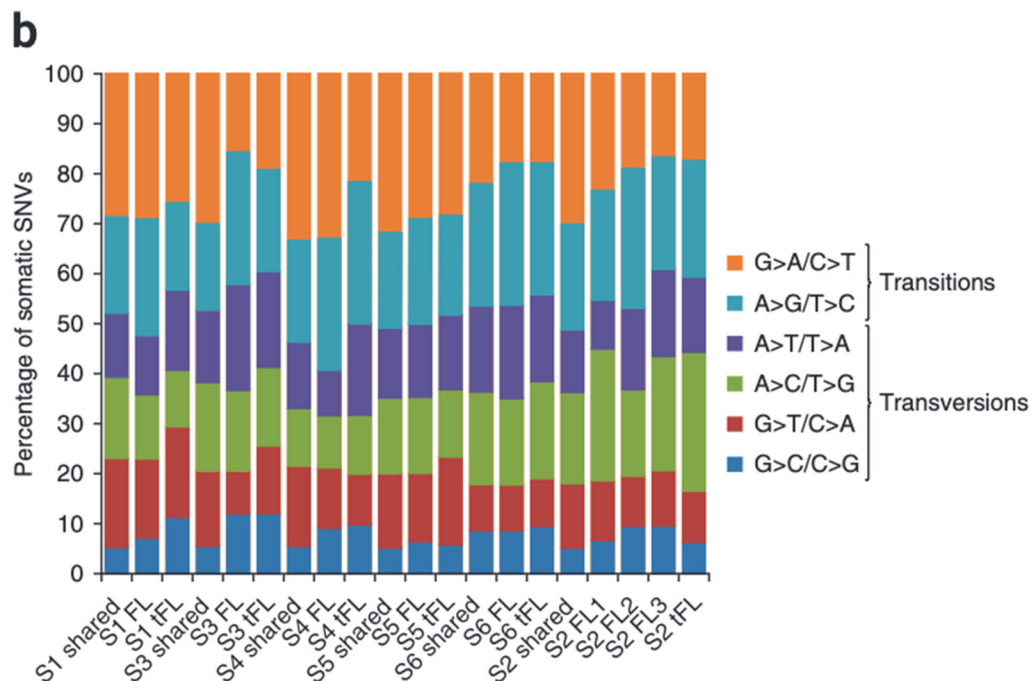
To determine the evolutionary relationship between the 10 paired follicular lymphoma-transformed follicular lymphoma samples, we constructed phylogenetic trees for each case using the somatic variants detected (only nonsynonymous variants and all variants). Each tree showed branching rather than linear evolution, with the ‘trunk’ representing genetic events shared by the follicular lymphoma tumor(s) and their paired transformed follicular lymphoma samples, thus supporting the presence of an ancestral CPC.

Id. at 177.

d) Step 1(d): “providing a report based on an identification of said mosaic variants specific to said tissue sample of said subject in said first additional sample, wherein said report comprises predicting, diagnosing, and/or prognosing a status or outcome of a cancer in said subject.”

253. Okosun-II provides a report based on an identification of the shared mosaic variants – i.e., the mosaic variants specific to the FL sample that were subsequently identified in the tFL sample. Ex. 1102 at 177 (Figure 1b). This report, at a minimum, diagnoses and/or prognoses the status of a cancer – namely and types (e.g. transitions and/or transversions) and the frequencies of mosaic mutations present in the tFL sample. *See* Ex. 1001 at 14:19-34 (“The methods may comprise reporting the number of reads, localization, cell type, spatial position in the body, and/or other values based on the comparison. . . The methods may comprise reporting the changes in the overall set of sequences.”).¹⁹

¹⁹ Okosun-I provides a report on disease status and likelihood of a response to “a stratified medicine approach targeting []specific pathways in FL” identified through its analysis. Ex. 1101. (analyzing “overlap of somatic variants (based on comparisons between non-synonymous, synonymous and UTR variants) between the paired FL/tFL biopsies” as well as the degree to which “this clonal relationship varied across the patients (range: 7.0–86.9%; mean 52.4%).”).



254. Thus, Okosun-II anticipates claim 1.

2. Claim 8

255. Claim 8 recites “[t]he method of claim 1, further comprising providing a recommendation for a therapeutic intervention based on an identification of said mosaic variants specific to said tissue sample of said subject.”

256. Okosun-II discloses that certain of the mosaic variants it detected that affect the NF-kB signalling pathway. Ex. 1102 at 178-179. Okosun-II provides a recommendation for a therapeutic intervention in cases where such mosaic variants are detected: “A number of emerging therapeutic agents target the NF-kB signaling cascade, and such treatments may be beneficial to a preselected subset of patients

with follicular lymphoma harboring mutations in this pathway” *Id.* at 179. *See also id.* at 176 (“An in-depth chronicling of the underlying genetic events from follicular lymphoma to transformed follicular lymphoma will serve as a guide in the development of effective targeted therapies.”); *id.* at 180 (“The high frequency of relapse suggests that CPCs are resistant to standard therapies and that adopting a stratified approach targeting the specific early genetic lesions identified in the putative CPC may ultimately offer the best chance of eradicating these cells and curing follicular lymphoma.”).

257. Thus, Okosun-II anticipates claim 8.

3. Claim 9

258. Claim 9 recites “[t]he method of claim 1, wherein said tissue sample comprises a metastatic tissue sample, a mixed tissue sample, a biopsy, or a core needle biopsy.”

259. As discussed above, Okosun-II utilized FL biopsies as its tissue samples. *See, e.g.*, Ex. 1102 at 177.

260. Thus, Okosun-II anticipates claim 9.

4. Claim 11

261. Claim 11 recites, “[t]he method of claim 1, wherein said mosaic variants specific to said subject are selected from the group consisting of: copy number variation, small insertions and deletions (inDels), single nucleotide

polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), and any combination thereof.”

262. The mosaic variants identified in Okosun-II include at least SNVs (i.e., SNPs and indels: “We verified 235 single-nucleotide variants (SNVs) and 48 indels out of 293 variants, corresponding to 118 genes.” *Id.* Okosun-II also discloses that “[i]nterrogation of the six whole-genome sequenced cases identified copy number alterations (CNAs), previously unreported structural variants, fusion transcripts and recurrent mutations in noncoding regions.” *Id.*

263. Thus, Okosun-II anticipates claim 11.

5. Claim 12

264. Claim 12 recites, “[t]he method of claim 1, wherein said predicting, diagnosing, and/or prognosing of (d) comprises one or more of the following: (i) diagnosing a cancer, (ii) identifying a cancer, (iii) determining a stage of a cancer, (iv) assessing risk of a cancer, (v) assessing risk of cancer recurrence, (vi) assessing reproductive risk, (vii) assessing efficacy of a drug, (viii) assessing risk of an adverse drug reaction, (ix) predicting optimal drug dosage, (x) predicting drug resistance, or any combination thereof.”

265. Okosun-II provides a report based on an identification of the shared mosaic variants – i.e., the mosaic variants specific to the FL sample that were subsequently identified in the tFL sample. Ex. 1102 at 177 (Figure 1b). This report,

at a minimum, diagnoses and/or prognoses the status, identity, and/or stage of a cancer, including based on the types (e.g. transitions and/or transversions) and the frequencies of mosaic mutations present in the tFL sample. *See* Ex. 1001 at 14:19-34 (“The methods may comprise reporting the number of reads, localization, cell type, spatial position in the body, and/or other values based on the comparison. . . The methods may comprise reporting the changes in the overall set of sequences.”).

266. Thus, Okosun-II anticipates claim 12.

6. Claim 13

267. Claim 13 recites, “[t]he method of claim 1, wherein said mosaic variants specific to said tissue sample of said subject are identified using capture probes in said first additional sample of step (c).”

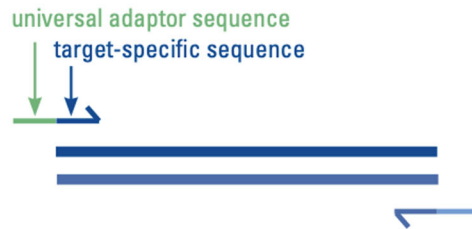
268. Okusun-II discloses identifying the tumor-specific mosaic variants identified in the initial whole-genome sequencing assay in a first additional sample using capture probes. Specifically, Okusun-II conducted “deep targeted resequencing of 28 genes in an extension cohort.” Ex. 1102 at 176. “To evaluate the mutation prevalence for the genes identified in our discovery cohort, we performed deep targeted resequencing of 28 candidate genes . . . in an extension cohort . . . including the 10 index cases, achieving a mean depth of coverage of 840x.” *Id.* at 177. The use of “targeted resequencing” was to specifically identify the tumor-

specific mosaic variants in additional samples. And as explained below, the system that Okosun-II used to do so used capture probes.

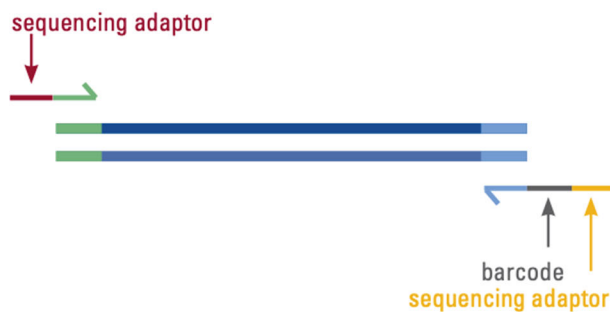
269. To perform its targeted resequencing, Okosun-II used “the Access Array platform (Fluidigm).” *Id.* at 182. A POSA would have been familiar with this system, and would have known that the Access Array System “is a unique target enrichment platform designed for resequencing selected regions from large numbers of samples.” Ex. 1110 (2012 Access Array Specification Sheet) at 1. The Access Array platform captures the sequences of interest using PCR primers as capture probes, and creates amplicons that contain further capture sequences appropriate for sequencer chemistry. *Id.* at 2. The manner in which Access Array Target Specific Primers capture sequences of interest is depicted in the schematic reproduced below.

Id.

- 1 Hybridization of sequence-specific primers to appropriate region of genomic DNA. Primers contain universal tag sequences to allow binding of barcode primers.



- 2 Hybridization of barcode primers, which also contain a capture sequence appropriate for sequencer chemistry



- 3 Final amplicon contains barcode sequence to identify parent DNA sample, and is tagged for capture and entry into emPCR.



Figure 1: Amplicon tagging, barcoding and sequencing

270. Accordingly, Okosun-II anticipates claim 13.

E. Ground 5: Claims 1-13 Would Have Been Obvious over Okosun-II in view of Knowledge of a POSA

271. As noted above, Okosun-II anticipates claim 1, which I incorporate by reference herein. Okosun-II in view of the knowledge of a POSA would have further rendered claims 1-13 obvious, as explained below.

1. Claim 1

272. Even had Okosun-II's use of peripheral blood as a source of germline DNA not anticipated the claimed use of leukocytes, a POSA would have found it obvious to do so. A POSA would have known that whole peripheral blood comprises leukocytes. Ex. 1109 (Qiagen Protocol) at 29 (stating that "[a] 200 μ l sample of whole human bloods" contains a concentration of $\sim 5 \times 10^6$ leukocytes/ml). Meyerson further notes that "[t]he samples generally used for germline genome analysis [are] peripheral blood mononuclear cells." Ex. 1031 at 686. Peripheral blood mononuclear cells, by Patent Owner's own admission, constitute leukocytes. Ex. 1119 (contending that Foresight's use of PBMCs in its methods satisfies the "leukocyte" limitation). Moreover, a POSA would have found it obvious to extract DNA from buffy coat, which is "a leukocyte-enriched fraction of whole blood," to obtain germline DNA for a sequencing experiment because "[p]reparing a buffy-coat fraction from whole blood is simple and yields approximately 5-10 times more DNA than an equivalent volume of whole blood." *Id.* Thus, the selection of a blood sample containing leukocytes as a source of germline DNA – whether whole blood, PBMCs, or buffy coat – would have been a routine and predictable experimental design choice in light of Okosun-II's express use of peripheral blood as a germline DNA sample.

2. Claim 2

273. Claim 2 recites, “[t]he method of claim 1, further comprising identifying mosaic variants specific to said tissue sample of said subject in a second additional sample of said subject, wherein said second additional sample comprises cell-free DNA, which second additional sample is obtained at a time point that is different than said tissue sample and said first additional sample.”

274. As explained above, Okosun-II anticipates and renders obvious claim 1. As explained with respect to step 1(c) above, Okosun-II discloses identifying tumor-specific mosaic variants serially to monitor the course of a given patient’s disease over time. *See, e.g.*, Ex. 1102 at 177 (Fig. 1); *id* at 178 (Fig. 2). To do so, Okosun-II discloses obtaining biopsies of the patient’s tumor at each time tested. *See* Fig. 1. Patient 2, for example, underwent surgery to resect lymph node biopsies at least 5 times, and all patients studied were biopsied at least 2 times. *Id.*

275. Patent Owner and its expert, Dr. Furneaux, admitted in co-pending proceedings involving patents asserted by Patent Owner against Petitioner that a POSA would have been motivated to avoid serial biopsies of the same patient. For example, Dr. Furneaux testified that “[a] POSA would have been dissuaded from obtaining a second tumor sample . . . because it would have involved an invasive procedure.” Ex. 1114 (Furneaux Declaration, Ex. 2031) at ¶136. Dr. Furneaux confirmed that a POSA would have been motivated to monitor tumor-specific

mosaic variants using non-invasive methods. *Id.* (Adopting Newman: “***Noninvasive access to malignant DNA is particularly attractive for solid tumors***, which cannot be repeatedly sampled without invasive procedures.”). Thus, though Okosun-II successfully performed its methods by “repeatedly samp[ling] with[] invasive procedures,” a POSA would have been motivated to obviate such repeated invasive procedures by gaining noninvasive access to the FL tumor (which is a solid tumor) DNA.

276. To that end, a POSA would have known that circulating tumor DNA in the plasma was a useful tool for noninvasively tracking tumor-associated mutations in blood samples taken over time. Moreover, those in the art would have known that as in other solid cancers, FL tumors released circulating tumor DNA into the blood. *See* Ex. 1111 (Hohaus) at 1412 (“Our study demonstrates that lymphoma patients frequently have elevated levels of cell-free circulating DNA at diagnosis, which can be accurately quantified by real-time PCR and which correlates with clinical parameters and prognosis.”); *see also* Section Section V.D, above.

277. Thus, a POSA would have found it obvious to apply, and reasonably would have expected success in applying, any one of the various known techniques for detecting tumor-specific variants in cfDNA from plasma, including those used in Chan, Dawson, Leary and McBride. Utilizing such techniques, and with knowledge of the sequences of the mosaic variants of interest based on its initial

WGS assay, a POSA would have found it obvious to utilize capture probes or plasma shotgun sequencing to noninvasively perform serial monitoring of the identified mosaic variants, as explained Section VIII.C.4 (Ground 3, Claim 2), which I incorporate by reference here.

278. **Diehn-530 (Ex. 1112):** Diehn-530 further evidences the knowledge of a POSA and discloses “[m]ethods for creating a selector of mutated genomic regions and for using the selector set to analyze genetic alterations in a cell-free nucleic acid sample. . . . The methods can be used to measure tumor-derived nucleic acids in a blood sample from a subject and thus to monitor the progression of disease in the subject. The methods can also be used for cancer screening, cancer diagnosis, cancer prognosis, and cancer therapy designation.” Abstract.

279. Diehn-530 explains that its methods “are provided for the highly sensitive analysis of circulating tumor DNA (ctDNA), e.g. DNA sequences present in the blood of an individual that are derived from tumor cells,” and that for its methods, “[t]umors of particular interest are solid tumors, including . . . lymphomas.” *Id.* at [0012]. In Table 10, Diehn-530 provides an exemplary selector (capture probe) set “comprising one or more genomic regions . . . useful for follicular lymphoma (FL)” ctDNA analysis. *Id.* at [0530].

280. Diehn-530 further explains methods for making personalized selector sets based on the identification of tumor-specific mosaic variants in paired versus

normal whole-genome sequencing, as in Okosun. *Id.* at [880] (“Here, tumor(s) from a patient known to have cancer are genotyped by profiling the tumor genome, exome, or targeted region expected to be enriched for somatic aberrations. The genotype of the cancer may be compared to a genotype of the germline of the same patient.”). Capture probes are then created. *Id.* (“The resulting lesions are then catalogued and used to build a custom, personalized selector comprising a set of biotinylated oligonucleotides for selective hybrid affinity capture of corresponding circulating tumor DNA (ctDNA) molecules.”). This personalized selector, i.e., a capture probe set, can be “applied for capture of the fragments of interest” in the cfDNA, “allowing the tracking and quantitation of those mutations originally discovered in the primary tumor within the corresponding cfDNA.” *Id.* In addition to the use of biotinylated oligonucleotide baits, Diehn-530 further explains that PCR-based capture probes may be used. *Id.* (“an alternative to affinity based hybrid capture of ctDNA/cfDNA, amplicons specific to the corresponding region could be interrogated by PCR[.]”).

281. Accordingly, Okosun-II in view of knowledge of a POSA renders claim 2 obvious.

3. Claim 3

282. Claim 3 recites, “[t]he method of claim 2, wherein said first and/or second additional sample comprises a blood sample, a saliva sample, a buccal swab, a urine sample, a semen sample, or a cerebrospinal fluid sample of said subject.”

283. As explained with respect to claim 2, above, the use of serial plasma samples as a source of cell-free tumor DNA comprising mosaic variants is rendered obvious by Okosun-II in view of the knowledge of a POSA.

284. Thus, Okosun-II in view of the knowledge of a POSA renders claim 3 obvious.

4. Claim 4

285. Claim 4 recites, “[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are selected from the group consisting of: copy number variation, small insertions and deletions (inDels), single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), and any combination thereof.”

286. As explained above with respect to claims 1 and 2, Okosun-II discloses identifying mosaic variants to the tumor tissue sample in the form of tumor specific single nucleotide variants, which are also known as single nucleotide polymorphisms. *See, e.g.*, Ex. 1102 at 177 (“In total, 1,560 protein-altering variants affecting 908 genes were detected across the 10 cases, comprising missense changes (84.8%), short indels (8.9%) and nonsense mutations (6.3%)”).

287. Thus, Okosun-II in view of the knowledge a POSA renders claim 4 obvious.

5. Claims 5 and 6

288. Claim 5 recites, “[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are causal or predisposing for a cancer.”

289. Claim 6 recites, “[t]he method of claim 2, wherein said mosaic variants specific to said tissue sample of said subject are non-causal for a cancer.”

290. The ‘968 patent does not disclose any information regarding the identification of a particular mosaic variant as “causal” or “non-causal,” beyond the tautology that a mosaic variant may be either causal or non causal. *See, e.g.*, Ex. 1001 at 12:15-17 (“The specific mosaic variants can be used as biomarkers . . . even if they are not causal for medical condition.”); *id.* at 13:38-40 (“In some cases, the tumor variants are causal variants (i.e., tumor-causing).”). The specification does not provide any criteria for when a variant should be considered “causal” or not.

291. At a minimum, the mosaic variants identified in Okosun-II and other references reflecting the knowledge of a POSA necessarily constitute either causal or non-causal mutations. For example, Okosun-II states that “[l]ongitudinal analyses identified early driver mutations in chromatin regulator genes . . . whereas mutations in EBF1 and regulators of NF-kB signaling . . . were gained at transformation.” Ex. 1102 at 176; *see also id.* at 179 (“Mutations present in nearly all tumor cells (clonal) would suggest early events and therefore represent initiating ‘driver’ genes.”). Thus, Okosun-II identified cancer-causing mosaic variants. Okosun’s analysis identifying

driver genes also resulted in the ruling out of other mosaic variants as non-causal of cancer. Ex. 1102 at 177 (“In total, 1,560 protein-altering variants affecting 908 genes were detected across the 10 cases, comprising missense changes (84.8%), short indels (8.9%) and nonsense mutations (6.3%)”).

292. Accordingly, Okosun-II in view of the knowledge of a POSA renders obvious detecting both causal and non-causal mosaic variants, and thus claims 5 and 6.

6. Claims 7, 10, and 13

293. Okosun anticipates claim 1, and renders obvious claim 2, as discussed above.

294. Claim 7 recites, “[t]he method of claim 2, further comprising enriching said extracted nucleic acid molecules of said tissue sample, blood sample, first additional sample, or second additional sample.”

295. Claim 10 recites, “[t]he method of claim 1, further comprising enriching said extracted nucleic acid molecules of said tissue sample, blood sample, or first additional sample.”

296. Claim 13 recites, “[t]he method of claim 1, wherein said mosaic variants specific to said tissue sample of said subject are identified using capture probes in said first additional sample of step (c).”

a) Obviousness Based on Serial Monitoring of ctDNA in Plasma

297. As explained above with respect to claim 2, a POSA would have found it obvious to identify tumor-specific mosaic variants in serially obtained blood samples using techniques that identify said mosaic variants using capture probes, including PCR-based and solution phase capture probes. Such methods further constitute enriching the extracted nucleic acid molecules of the serially obtained additional samples.

b) Obviousness Based on Use of Whole Exome Sequencing for Additional Samples

298. In addition, Okusun-II renders obvious the use of capture probes to identify tumor-specific mosaic variants in an additional sample using whole exome sequencing. Specifically, Okusun-II explained that it “performed whole-genome or whole-exome sequencing on 10 follicular lymphoma-transformed follicular lymphoma pairs followed by deep sequencing of 28 genes in an extension cohort.” Ex 1102 at 176. Okusun’s findings and results would have motivated a POSA to perform less sequencing than whole-genome sequencing on additional serial samples in some applications. For example, Okusun demonstrated the success of whole exome sequencing in parallel with its WGS experiment, and such WES would have represented a reliable and predictable way to reduce the amount of sequencing required in subsequent samples after identifying mutations of interest via initial

WGS. The performance of such whole exome sequencing involves the use of capture probes.

299. For example, Okosun-II explains that after identifying mosaic variants using the WGS and WES data generated in its ten index cases, they “focus[ed] on protein-altering changes” and “protein altering variants” affecting the identified mosaic genes. *Id.* at 176-177. When focused specifically on tumor-specific mosaic variants that effect protein-encoding genes, a POSA would have been motivated to target sequencing to protein-encoding genes, and applying whole exome sequencing to subsequent samples would have represented an obvious upper bound on such sequencing. The exome refers to the portion of the genome that encodes genes. Thus, a POSA performing longitudinal analysis of protein-encoding genes over time, as disclosed and suggested by Okosun, would have been motivated to restrict the downstream analysis in its workflow to no more than the exome. Sequencing and analyzing “protein-altering changes” in subsequent samples using whole exome sequencing, as compared to WGS, would have been an economical approach to monitoring the mosaic variants identified in the initial WGS over time. A POSA would have reasonably expected success in doing so, given that Okosun-II reported successfully analyzing samples using both methods. This confirms that the choice of WES or WGS in its workflow involves only the predictable substitution of one known method for another.

300. Okosun-II further establishes that its whole exome sequencing involved the use of capture probes. The whole exome sequencing “[c]apture libraries were constructed from 3 µg of tumor and germline DNA using the Agilent SureSelect XT Human All Exon V4 kit.” It is well known in the art that the Agilent SureSelect product constituted a type of capture probes. *See* Ex. 1115 at ¶93 (Mr. West admits: “At least Mark Pratt and I were aware that ... capture probes could be purchased via Agilent (their ‘SureSelect’ line of capture probes)”).

301. Thus, Okosun-II in view of the knowledge of a POSA renders claim 13 obvious.

IX. Secondary Considerations of Nonobviousness

302. I understand that in previous IPRs, Personalis argued that its claimed inventions were the subject of industry praise on the basis of Personalis’s Next Personal assay. For example, in IPR2023-00224, Dr. Furneaux contends that “Personalis’ assay can detect tumor in DNA in blood plasma at levels of about one part per million,” whereas the prior art at issue allegedly was “several hundred times less sensitive than the claimed method.” *See, e.g.,* Ex. 1114 at ¶ 230. I note however, that none of the claims at issue recite a particular sensitivity, nor does Dr. Furneaux provide any reason why the purported sensitivity of Personalis’s assay is attributable to the claimed features, as opposed to other technology present in Next Personal.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: November 22, 2023

By: 
John Quackenbush, Ph.D.