

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

FORESIGHT DIAGNOSTICS INC.,
Petitioner,

v.

PERSONALIS, INC.,
Patent Owner.

IPR2023-00546
U.S. Patent No. 10,450,611

**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 '611 Patent.....	12
IV. Person of Ordinary Skill in the Art	13
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	23
1. Sequencing Techniques	23
2. Sequencing Depth and Read Counts	28
3. Amplification and Enrichment	32
D. Personalized Cancer Testing and Monitoring.....	41
E. Related Applications of Next Generation Sequencing	46
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. Ley (Ex. 1012)	51
B. Newman (Ex. 1043)	51
VII. Claim Construction.....	51
A. “probe”	53
B. “report”	54
VIII. Prior Art Anticipated or Would Have Rendered Obvious Claims 1-5, 9-11, 13, 18, 19, 28-30 of the '611 Patent	55
A. Ground 1: Ley Anticipates Claims 1, 2, 4, 5, 9-11, 13, 18, 19, 28-30.....	55
1. Claim 1	57
2. Claim 2	69
3. Claim 4	73
4. Claim 5	74
5. Claim 9	76
6. Claim 10	77
7. Claim 11	79
8. Claim 13	79
9. Claim 18	80
10. Claim 19	80
11. Claim 28	81
12. Claim 29	83

TABLE OF CONTENTS
Cont'd

	<u>Page</u>
13. Claim 30.....	83
B. Ground 2: Ley Renders Obvious Claims 1-5, 9-11, 13, 18, 19, 28-30	84
1. Claims 1-5, 9-11, 13, 18, 19, 28-30	85
2. Claim 3—Further Information Regarding Obviousness	87
3. Claim 5—Further Information Regarding Obviousness	88
4. Claim 28—Further Information Concerning Obviousness	92
Ground 3: Newman Anticipates Claims 1-5, 9-11, 13, 18, 19, 28-30.....	93
5. Claim 1	95
6. Claim 2.....	104
7. Claim 3.....	105
8. Claim 4.....	105
9. Claim 5.....	106
10. Claim 9.....	107
11. Claim 10.....	108
12. Claim 11.....	108
13. Claim 13.....	110
14. Claim 18.....	111

TABLE OF CONTENTS
Cont'd

	<u>Page</u>
15. Claim 19.....	111
16. Claim 28.....	112
17. Claim 29.....	112
18. Claim 30.....	112
C. Ground 4: Newman Renders Obvious Claims 1-5, 9-11, 13, 18, 19, 28-30	113
1. Claims 1-5, 9-11, 13, 18, 19, 28-30	114
2. Claim 5—Further Information Regarding Obviousness	116

TABLE OF AUTHORITIES

	Page(s)
Statutes	
35 U.S.C. § 103(a)	9
United States Code Title 18 Section 1001	122
Other Authorities	
European Nucleotide Archive (https://www.ebi.ac.uk/ena).....	31
(https://www.garvan.org.au/research/kinghorn-centre-for-clinical-genomics/learn-about-genomics/for-gp/genetics-refresher-1/types-of-variants).....	21
(https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data)	26
<u>https://www.genome.gov/about-genomics/fact-sheets/Polymerase-Chain-Reaction-Fact-Sheet</u> (last accessed 11/15/2022)	38

Exhibits Cited

Ex. 1001	U.S. Patent No. 10,450,611 issued to West <i>et al.</i>
Ex. 1002	Rebecca J. Leary <i>et al.</i> , “Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing,” <i>Science Translational Medicine</i> (2010) (Leary)
Ex. 1003	Quake <i>et al.</i> , WO 2011/057061 (Quake)
Ex. 1004	Unused
Ex. 1005	Andreas Gnirke <i>et al.</i> , “Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing,” <i>Nature Biotechnology</i> (2009) (Gnirke)
Ex. 1006	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)
Ex. 1007	Lira Mamanova <i>et al.</i> , “Target-enrichment strategies for next-generation sequencing,” <i>Nature</i> (2010) (Mamanova)
Ex. 1008	Unused
Ex. 1009	Unused
Ex. 1010	<i>Personalis v. Foresight</i> , First Amended Complaint, D. Colo., Case No. 1:22-v-01913, Dkt. 9, filed August, 17, 2022 (“FAC”)
Ex. 1011	Unused
Ex. 1012	Timothy J. Ley <i>et al.</i> , “DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome,” <i>Nature</i> (2008) (Ley)
Ex. 1013	Jay Shendure & Hanlee Ji, “Next-generation DNA sequencing,” <i>Nature Biotechnology</i> (2008) (Shendure)
Ex. 1014	Jeffrey S. Ross and Maureen Cronin, “Whole Cancer Genome Sequencing by Next-Generation Methods,” <i>American Journal of Clinical Pathology</i> (2011) (Ross)

Ex. 1015	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)
Ex. 1016	Unused
Ex. 1017	Unused
Ex. 1018	Unused
Ex. 1019	https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
Ex. 1020	Unused
Ex. 1021	Unused
Ex. 1022	Unused
Ex. 1023	<i>Curriculum Vitae</i> of John Quackenbush
Ex. 1024	https://www.garvan.org.au/research/kinghorn-centre-for-clinical-genomics/learn-about-genomics/for-gp/genetics-refresher-1/types-of-variants
Ex. 1025	Unused
Ex. 1026	Unused
Ex. 1027	Hugo Lam, et al., “Performance comparison of whole-genome sequencing platforms,” <i>Nature Biotechnology</i> (2012) (Lam)
Ex. 1028	Illumina, Technical Note: Sequencing
Ex. 1029	Illumina, Specification Sheet: Illumina Sequencing
Ex. 1030	Unused
Ex. 1031	Matthew Meyerson, et al., “Advances in understanding cancer genomes through second-generation sequencing,” <i>Nature Reviews</i> (2010) (Meyerson)

Ex. 1032	Unused
Ex. 1033	Unused
Ex. 1034	Unused
Ex. 1035	Unused
Ex. 1036	Unused
Ex. 1037	Infringement Claim Chart for U.S. Patent No. 10,450,611; Ex. D to Patent Owner's FAC
Ex. 1038	Unused
Ex. 1039	Unused
Ex. 1040	D. Sims <i>et al.</i> , "Sequencing depth and coverage: key considerations in genomic analyses," Nature Reviews: Genetics (2014) (<i>Sims</i>)
Ex. 1041	H. Li, "Mapping short DNA sequencing reads and calling variants using mapping quality scores," Genome Res. (2008) (<i>Li</i>)
Ex. 1042	Ley Supplemental Information
Ex. 1043	A. Newman <i>et al.</i> , "Integrated digital error suppression for improved detection of circulating tumor DNA," Nature Biotechnology (2016) (<i>Newman</i>)
Ex. 1044	A. Newman <i>et al.</i> , "An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage," Nature Medicine (2014) (<i>Newman 2014</i>)
Ex. 1045	Newman Supplementary Excel Spreadsheet
Ex. 1046	M. Clark <i>et al.</i> , "Performance comparison of exome DNA sequencing technologies," Nature Biotechnology (2011) (<i>Clark</i>)
Ex. 1047	Illumina MiSeq Worksheet (2015)
Ex. 1048	Genbank Record for CDH24

Ex. 1049	Genbank Record for SLC15A1
Ex. 1050	Genbank Record for KNDC1
Ex. 1051	Genbank Record for PTPRT
Ex. 1052	Genbank Record for GRINL1B
Ex. 1053	Genbank Record for GPR123
Ex. 1054	Genbank Record for EBI2
Ex. 1055	Genbank Record for PCLKC
Ex. 1056	Genbank Record for FLT3
Ex. 1057	Genbank Record for NPM1
Ex. 1058	Genbank Record for BRCA2
Ex. 1059	Genbank Record for TP53
Ex. 1060	Newman Declaration
Ex. 1061	Quackenbush Declaration
Ex. 1062	U.S. Patent No. 11,299,783, issued to West <i>et. al.</i>
Ex. 1063	Unused
Ex. 1064	Unused
Ex. 1065	Unused
Ex. 1066	Unused
Ex. 1067	Unused
Ex. 1068	Unused
Ex. 1069	Unused
Ex. 1070	Unused

Ex. 1071	R. Burrell, et al., “The causes and consequences of genetic heterogeneity in cancer evolution,” Nature (2013) (<i>Burrell</i>)
Ex. 1072	Unused
Ex. 1073	C. Haferlach, et al., “Mutations of the TP53 gene in acute myeloid leukemia are strongly associated with a complex aberrant karyotype,” Leukemia (2008)
Ex. 1074	S. Misawa, et al., “Significance of chromosomal alterations and mutations of the N-RAS and TP53 genes in relation to leukemogenesis of acute myeloid leukemia,” Leukemia Research (1998) (<i>Misawa</i>)
Ex. 1075	S. Davies, et al., “Indications for Hematopoietic Cell Transplantation in Acute Leukemia,” Biology of Blood and Marrow Transplantation (2008) (<i>Davies</i>)
Ex. 1076	Unused
Ex. 1077	D. Bentley, et al., “Accurate whole human genome sequencing using reversible terminator chemistry,” Nature (2008) (<i>Bentley</i>)

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. 10,450,611 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 ’611 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. Exs. 1010, 1037.

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; and (3) the level of ordinary skill in the pertinent art.

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 '611 Patent

33. The '611 patent (Ex. 1001) is titled "Personalized Genetic Testing." Ex. 1001, Code (54). The patent claims priority to "[p]rovisional application No. 62/342,674, filed on May 27, 2016." *Id.*, Code (60). I have considered May 27, 2016 to be the relevant date for assessing prior art invalidity in my analysis, although I offer no opinion on whether the '674 provisional application in fact provides sufficient support for the issued claims of the '611 patent.¹

34. In its complaint for patent infringement against Petitioner, Patent Owner characterized the subject matter claimed in the '611 patent as follows:

For example, claim 1 of the '611 Patent claims a method for personalized genetic testing comprising of sequencing nucleic acid molecules from a biological sample, identifying nucleic acid sequences having a set of genetic variants, obtaining a set of nucleic acid probe molecules configured to selectively enrich or amplify sequencing with genetic variants, using the set of nucleic acid probe molecules to selectively enrich or amplify sequences with genetic variants in another biological sample to generate a sequencing library, and sequencing the

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

sequencing library to identify presence or absence of at least some genetic variants.

Ex. 1010 (Personalis First Amended Complaint) at ¶18.

IV. Person of Ordinary Skill in the Art

35. 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 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.

36. The relevant technology field for the ’611 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 ’611 patent before May 27, 2016.

37. After considering the factors above, I believe a person of ordinary skill in the relevant field before May 27, 2016 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.

38. 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

39. In this Section, I provide an overview of the knowledge that a POSA would have possessed as of May 27, 2016.

A. Overview of Molecular Biology and Genetics

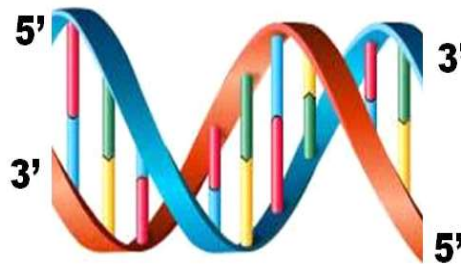
40. I begin with an overview of the principles of biology that are relevant to understanding the state of the art, especially those pertaining to nucleic acid sequencing and cancer biology.

1. The Central Dogma of Molecular Biology

41. 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.

42. 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.

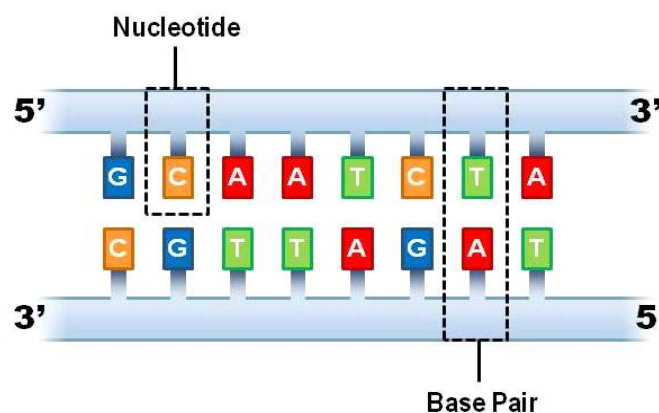


43. The smallest functional unit of DNA is the gene, which is a sequence of nucleotides that encode for ribonucleic acid (“RNA”), which in turn codes for a protein. Proteins are incredibly versatile large molecules that serve crucial functions in essentially all biological processes. As examples: some 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.

44. 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.”

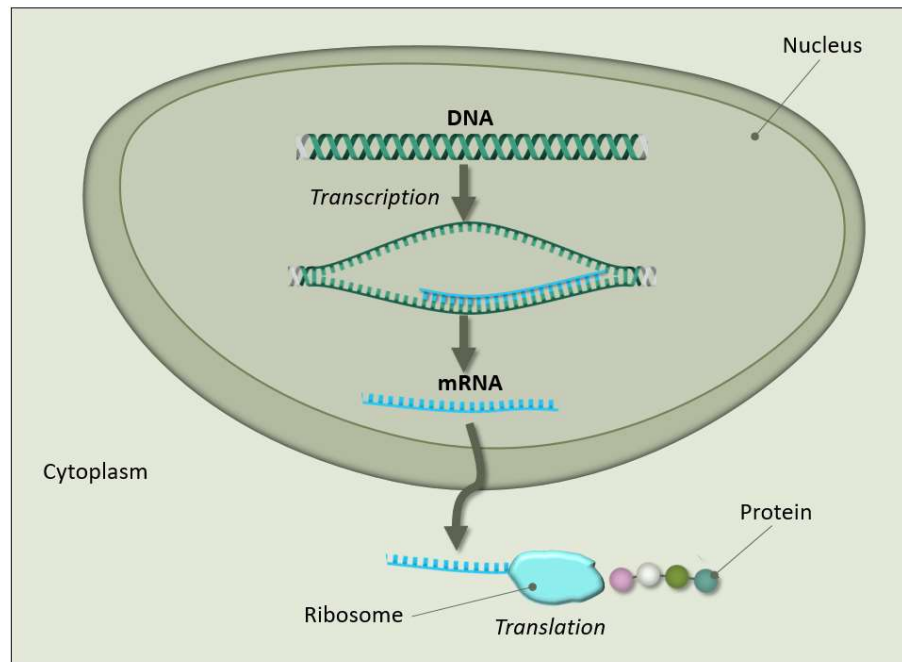
45. There are four natural types of bases that occur in DNA: adenine (“A”), cytosine (“C”), guanine (“G”) and thymine (“T”). Each type of base on one DNA strand binds to only one type of base on the other strand, known as its “complementary” base. When complementary bases bind together, or “hybridize,” 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.



46. 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.

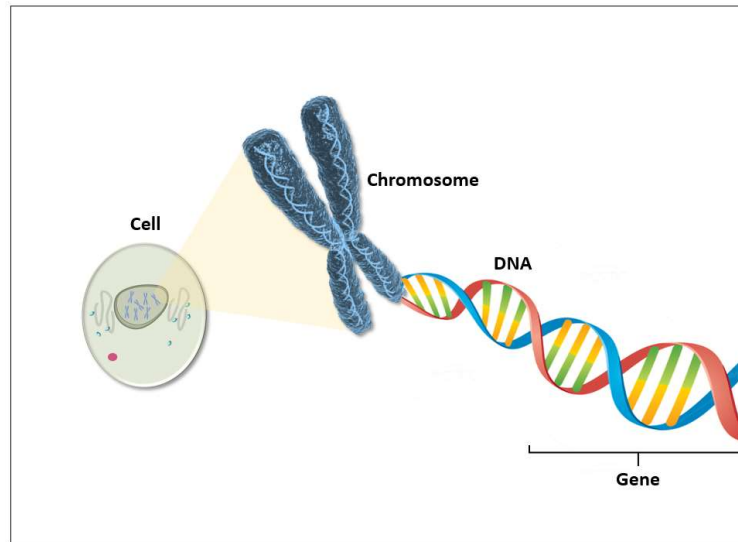
47. The information encoded by the DNA in a gene is turned into a protein via two step-wise processes, called transcription and translation. In transcription, the DNA is used as a template by specialized machinery in the cell to generate another type of molecule, called messenger ribonucleic acid (mRNA, 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.

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



2. Genetics

49. 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.



50. 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."

51. 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.

52. 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.

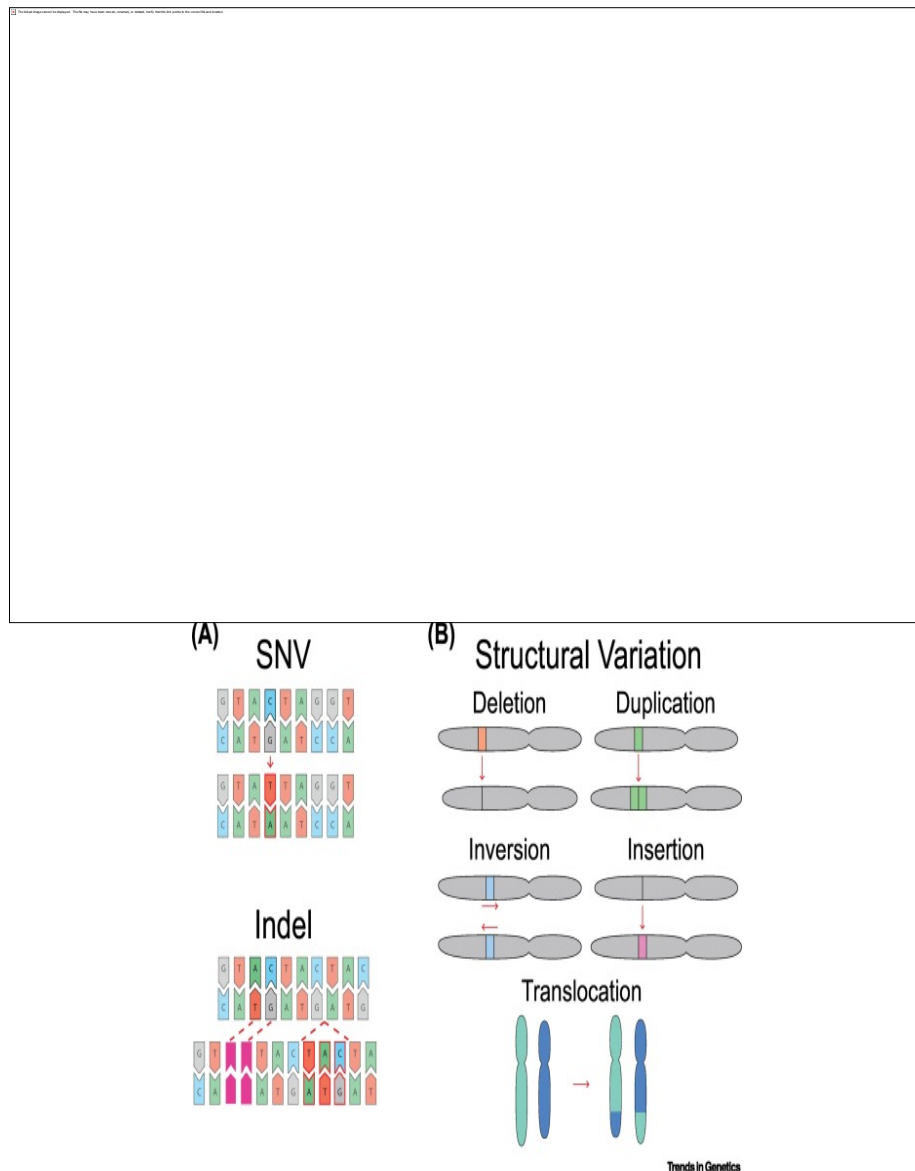
B. Genetic Mutations

53. 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.

54. Different types of somatic mutations were known in the prior art well before May 2016. *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.



55. 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.

56. By 2016, 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”; the ICGC Data Portal), The Cancer Genome Atlas (“TCGA”; the Genomic Data Common Portal), International Agency for Research on Cancer (“IARC”; *e.g.* the IARC/WHO TP53 database), and the Catalogue of Somatic Mutations in Cancer (“COSMIC”).

C. Nucleic Acid Sequencing Technologies

1. Sequencing Techniques

57. 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.

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

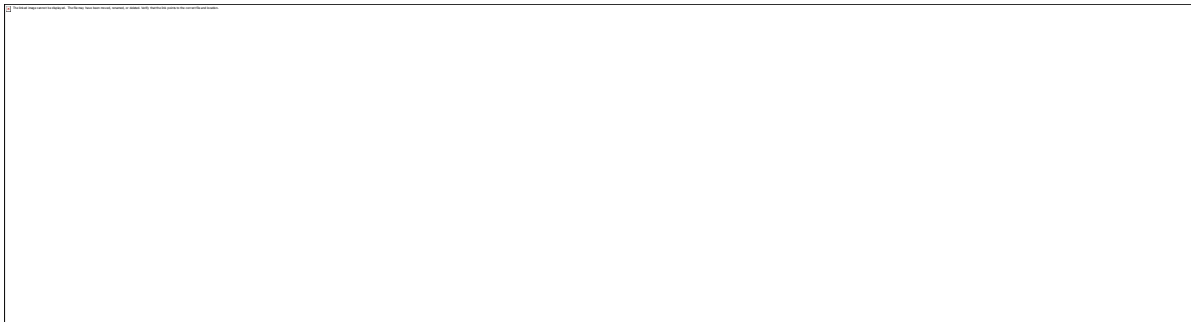
DNA, and incorporating labeled, chain-terminating bases at different locations along the different growing strands. The DNA fragments are then separated by size using electrophoresis and the various lengths of the fragments are used with the terminating bases to reconstruct the whole sequence of the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

59. 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.

60. 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

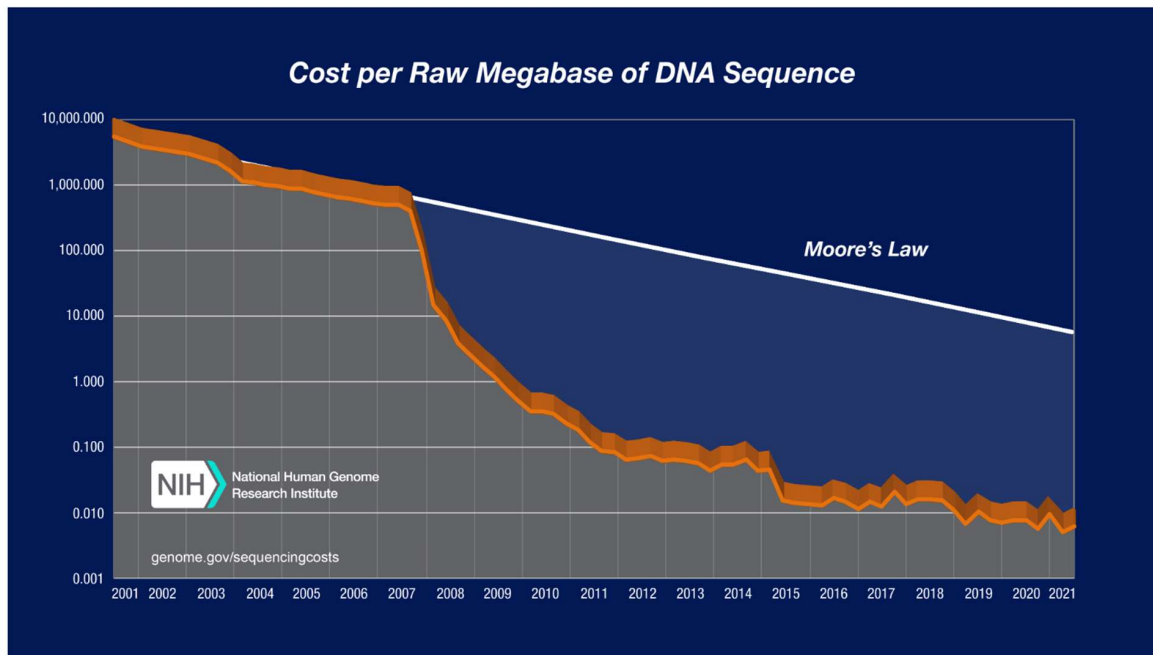
a di-deoxynucleotide incorporated randomly, leading to fluorescently labeled strands of different lengths.

61. 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.



62. 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). The cost of

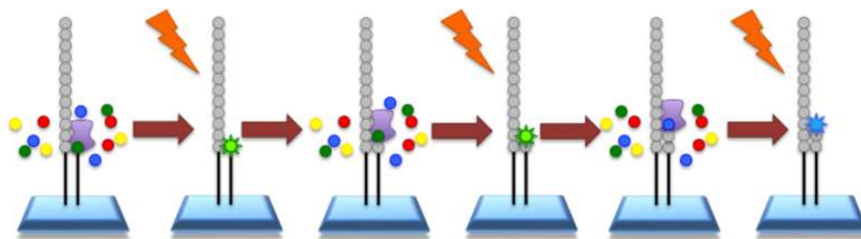
implementing NGS methods rapidly decreased, making such approaches more accessible and more desirable for further development. Ex. 1019 (<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>).

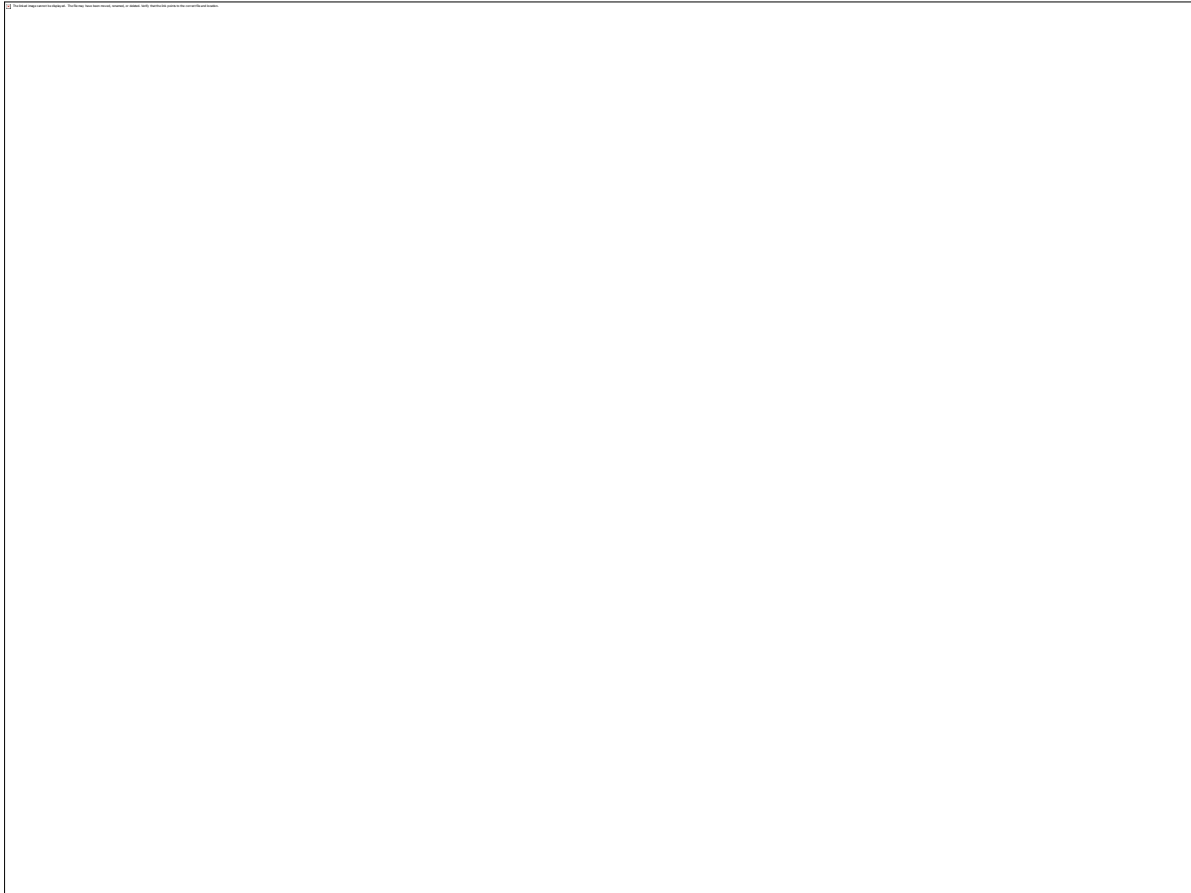


Sequencing cost per megabase - 2021

63. 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.





64. 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

65. NGS has been applied widely to sequencing of genomic sequences. It was known well before May 2016 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”).

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

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

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 sequencing 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 in many applications 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.

67. 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.

68. 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.

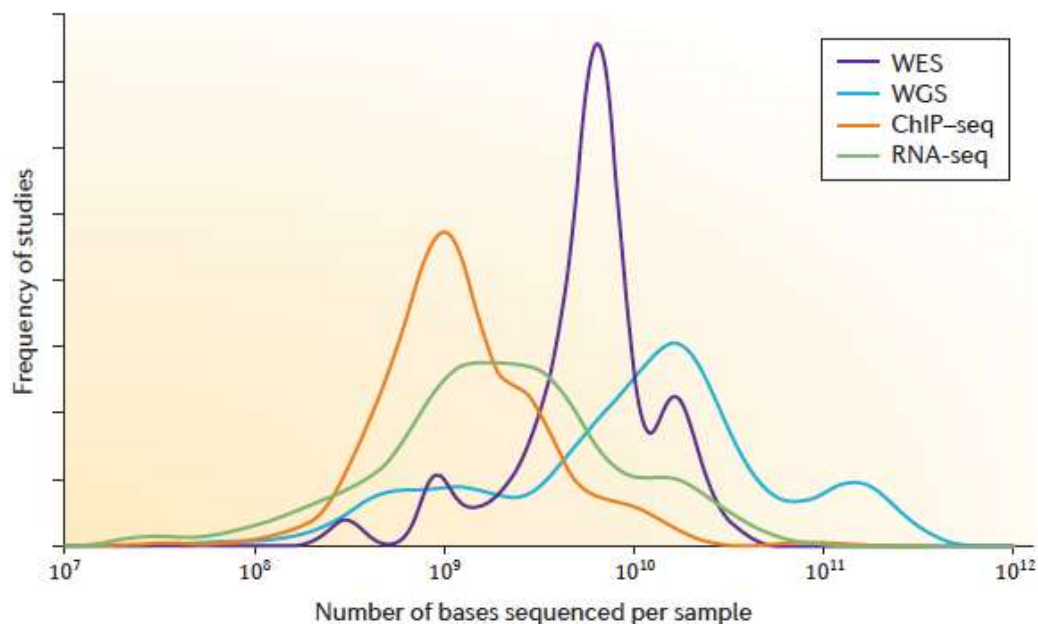


Figure 1 | Sequencing depths for different applications. The frequency of studies that use read counts of all runs (which are typically flow-cell lanes) and that were deposited from 2012 to June 2013 for the Illumina platform in the European Nucleotide Archive (ENA) is shown. The plot provides an overview 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

69. Both Sanger sequencing and NGS require minimum quantities and/or relative abundances of target DNA. Thus, well before 2016 several techniques had

been developed to generate or isolate copies of DNA molecules of interest to facilitate sequencing.

70. 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.

71. 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).

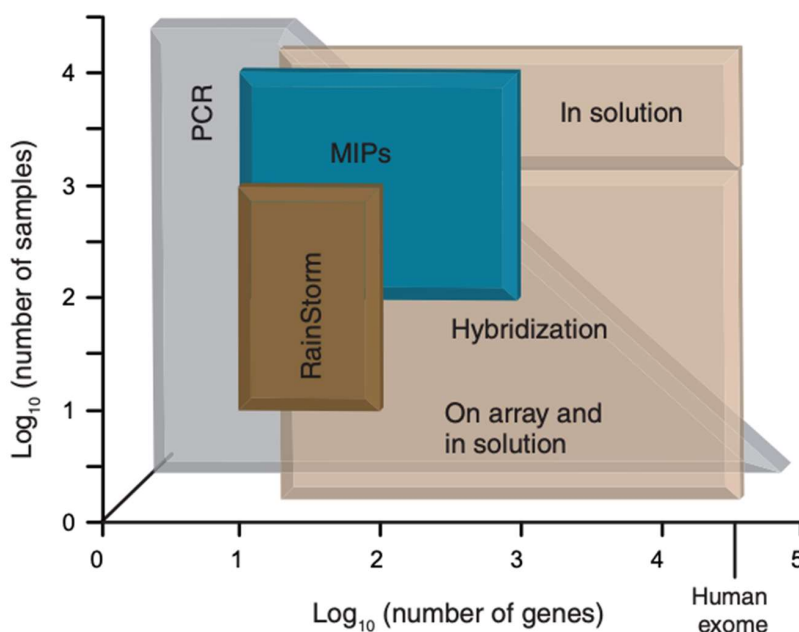


Figure 2 | Suitability of different target-enrichment strategies to different combinations of target size and sample number. Suitability was estimated from the perspective of the feasibility with which each method could be applied to the various combinations of target size and sample number, rather than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

72. 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 – 10^5 (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 SequaPrep, 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	10^2 – 10^4 (up to 10 Mb)	

All major targeted enrichment techniques show relative pros and cons.

a) PCR

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

74. 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.

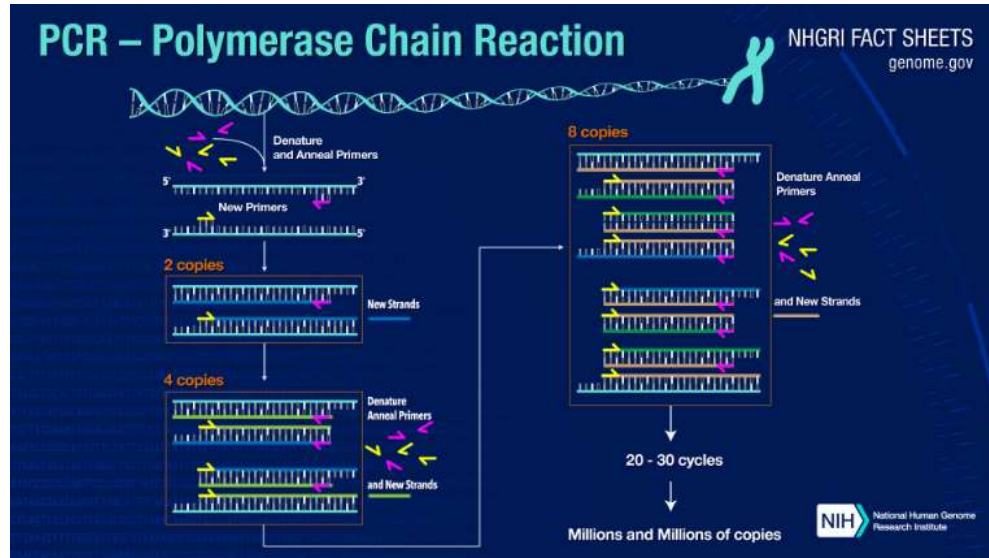
75. PCR “primers” or “probes,” which are short oligonucleotides that are complementary to a small portion of the DNA templates, can then be added to the denatured DNA at high concentrations. These primers/probes then hybridize (*i.e.*, bind through base pairing interactions) to the DNA template strand 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.”

76. 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.

77. 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.”

78. 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.²



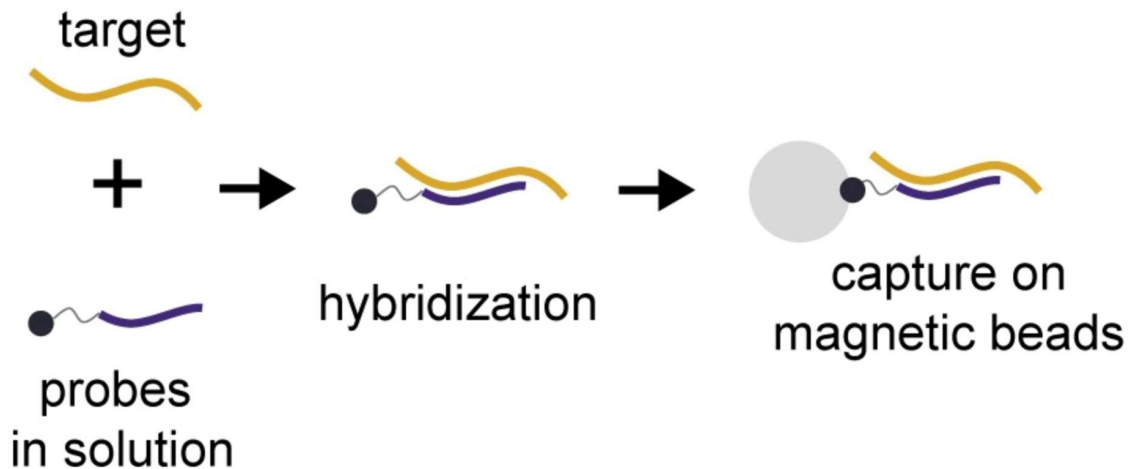
b) Solution-Based Hybrid Capture

79. 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)³

² See <https://www.genome.gov/about-genomics/fact-sheets/Polymerase-Chain-Reaction-Fact-Sheet> (last accessed 11/15/2022).

³ The streptavidin-biotin complex is well-known to be one of the strongest binding interactions occurring in nature. The binding strength is significantly greater than that between two complementary DNA strands, and is far beyond the normal protein-ligand binding strength.

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).

80. 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 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.”).

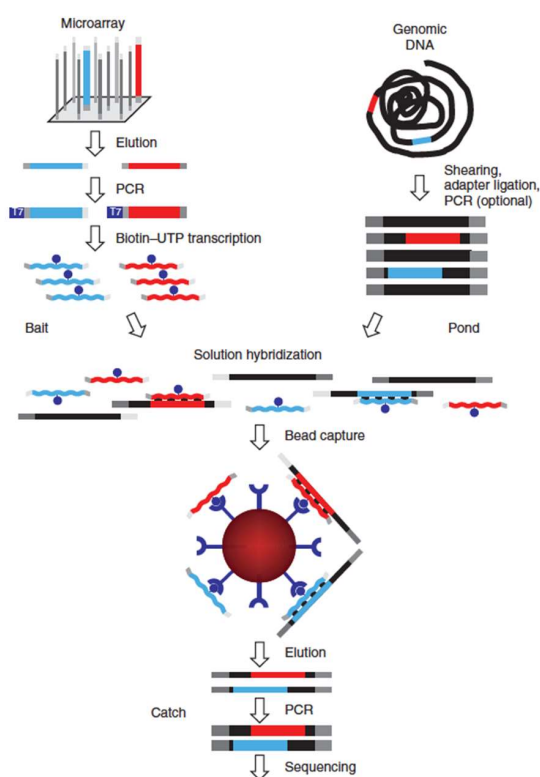


Figure 1 Overview of hybrid selection method. Illustrated are steps involved in the preparation of a complex pool of biotinylated RNA capture probes (bait; top left), whole-genome fragment input library (pond; top right) and hybrid-selected enriched output library (catch; bottom). Two sequencing targets 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.

81. 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 research 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

82. 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.

83. 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.*

84. 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 FLX^[4] 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 10 somatic and 2 germline (inherited) mutations that were analyzed at different stages of the disease (red and green bars, where somatic mutations should primarily be seen), and

⁴ 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.

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.

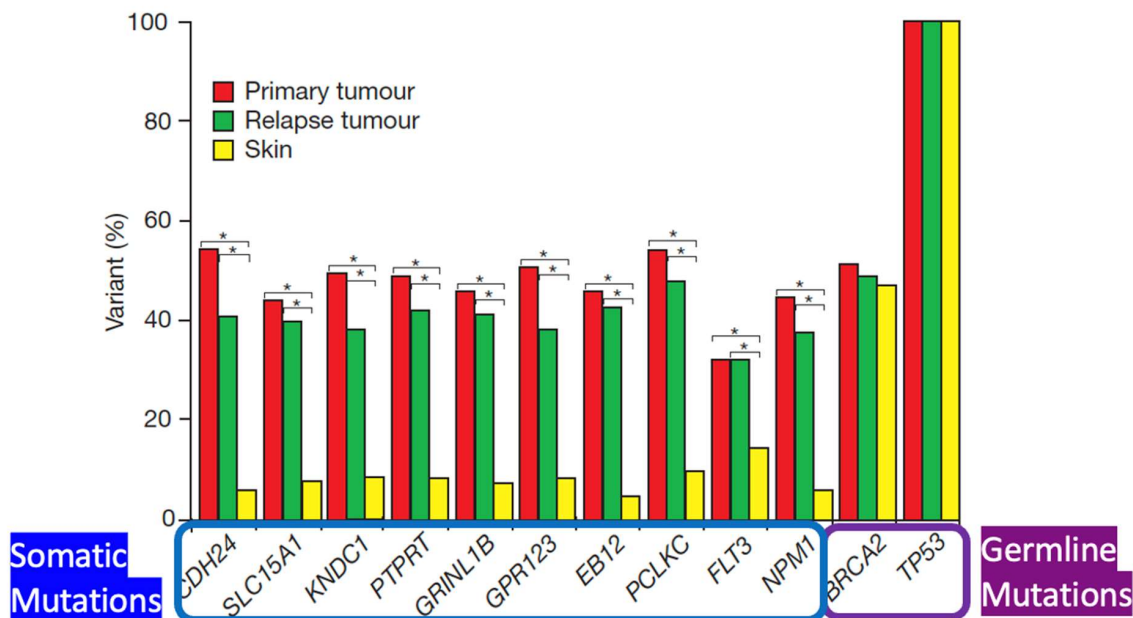


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that 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.

85. 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).

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

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)

(reproduced below). 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. Ley (Ex. 1012)

95. Ley *et al.*, “DNA sequencing of a cytogenetically normal acute myeloid leukemia genome,” is an article that was published in *Nature* by November 6, 2008, and is thus prior art to the ’611 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.

B. Newman (Ex. 1043)

96. Newman *et al.*, “Integrated digital error suppression for improved detection of circulating tumor DNA,” is an article that was published in *Nature Biotechnology* by May 2016. It was published online on March 28, 2016. Ex. 1060.

VII. Claim Construction

97. I have generally applied the plain and ordinary meaning that the claims of the ’611 patent would have had to a POSA in view of the ’611 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. Each of these terms (indicated in bold italics below) is present in claim 1-2 of the '611 patent:

Claim 1: A method for personalized genetic testing, comprising:

(a) sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more;

(b) computer processing said sequencing data to identify a plurality of nucleic acid sequences having a set of genetic variants, wherein said set of genetic variants are identified with respect to a reference;

(c) obtaining a ***probe*** set comprising a plurality of nucleic acid ***probe*** molecules having said plurality of nucleic acid sequences or complements thereof, wherein said plurality of nucleic acid ***probe*** molecules are configured to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample;

(d) using said ***probe*** set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in an additional biological sample from said subject or at least one biological relative of said subject, to generate a sequencing library; and

(e) subjecting said sequencing library to sequencing at a coverage of 175 kilobases (kb) or less to identify a presence or absence of at least a subset of said set of genetic variants in said additional biological sample from said subject or said at least one biological relative, wherein a

sequencing depth of said sequencing of (e) is greater than a sequencing depth of said sequencing of (a).

Claim 2: “[t]he method of claim 1, further comprising outputting a *report* that is generated at least based on comparison of results from said sequencing of (a) with results from at least said subjecting said sequencing library to said sequencing of (e).”

A. “probe”

98. I have applied the meaning the term “probe” would have had to a POSA in view of the ’611 patent’s specification. The ’611 patent explains, for example, that probes can be used in a hybridization array, by “synthesizing the plurality of nucleic acid probe molecules using at least one array” Ex. 1001 at 4:54-55; see also *id.* at 7:18-20 (probes include “synthesizing the plurality of nucleic acid probe molecules on a single solid substrate.”). The ’611 patent also explains that “nucleic acid probe molecules comprise primers for amplifying the nucleic acid sequences.” *Id.* at 5:16-18. And again, the ’611 patent states that “the plurality of nucleic acid probe molecules of the assay enrich or deplete a nucleic acid mixture of nucleic acid molecules for target regions, by hybridization or amplification.” *Id.* at 6:64-67. Thus, a POSA reading the specification of the ’611 patent would have understood that the claimed “nucleic acid probe molecules” include PCR primers, hybrid capture baits, and other such nucleic acid probe molecules for selectively amplifying or enriching sequences.

B. “report”

99. I have applied the meaning of the term “report” would have had to a POSA in view of the ’611 patent’s specification. Figure 1 of the ’611 patent, for example, is consistent with the plain meaning. It shows a therapeutic intervention, which can be a report, at the end of a process that includes personalized assays:

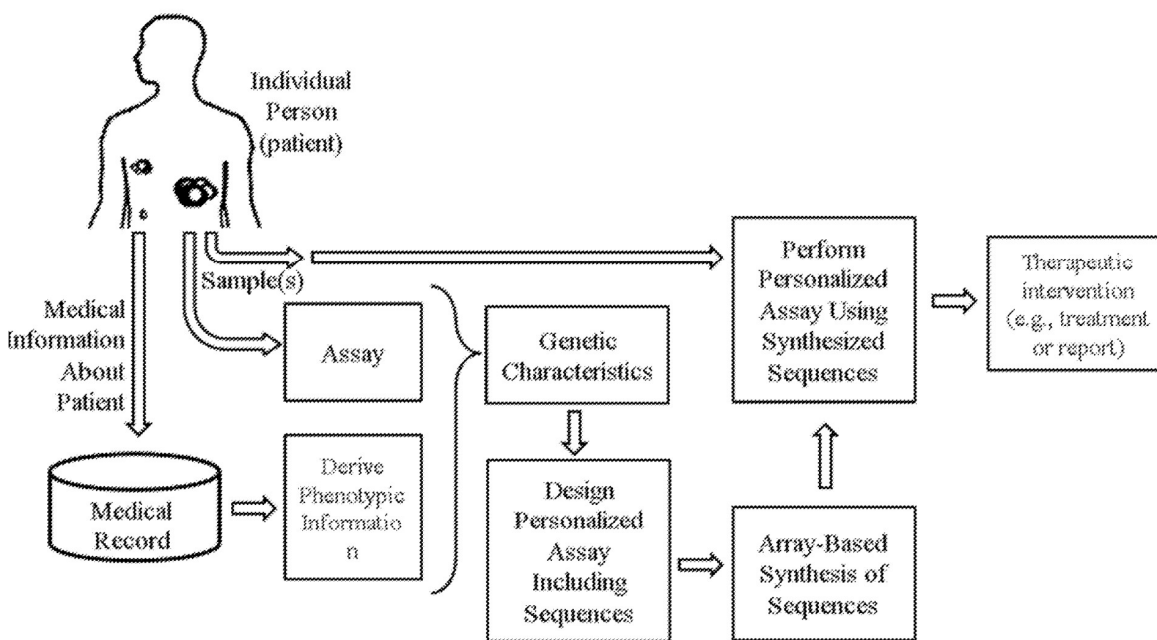


FIG. 1

100. The ’611 patent also explains that “[s]ome embodiments comprise outputting a report that is indicative of a presence or absence of the at least the subset of the genetic variants in the subject or the at least one biological relative.” Ex. 1001 at 5:13-16. Thus, a POSA reading the specification of the ’611 patent would have

understood that the claimed “report” includes, for example, providing information for treatment of a patient as well as providing an indication of genetic variation.

VIII. Prior Art Anticipated or Would Have Rendered Obvious Claims 1-5, 9-11, 13, 18, 19, 28-30 of the '611 Patent

A. Ground 1: Ley Anticipates Claims 1, 2, 4, 5, 9-11, 13, 18, 19, 28-30

101. Claim 1 recites “A method for personalized genetic testing, comprising” enumerated steps (a)-(e). As discussed in detail below, Ley discloses such a method, and thus anticipates claim 1.

102. In full, Claim 1 recites:

A method for personalized genetic testing, comprising:

(a) sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more;

(b) computer processing said sequencing data to identify a plurality of nucleic acid sequences having a set of genetic variants, wherein said set of genetic variants are identified with respect to a reference;

(c) obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or complements thereof, wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample;

(d) using said probe set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in an additional biological sample from said subject or at least one biological relative of said subject, to generate a sequencing library; and

(e) subjecting said sequencing library to sequencing at a coverage of 175 kilobases (kb) or less to identify a presence or absence of at least a subset of said set of genetic variants in said additional biological sample from said subject or said at least one biological relative, wherein a sequencing depth of said sequencing of (e) is greater than a sequencing depth of said sequencing of (a).

103. Ley discloses a method of genetic testing that used “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. The genetic testing described in Ley was “personalized” given that the assay sequenced the patient’s tumor’s genome and compared “its matched normal counterpart obtained from the same patient’s skin.” *Id.* Through this personalized genetic testing, 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 method of personalized genetic testing disclosed by Ley meets each limitation of claim 1, as discussed below.

1. Claim 1

104. The preamble of claim 1 recites “[a] method for personalized genetic testing.” Ley discloses such a method. The genetic testing described in Ley was “personalized” given that the assay sequenced the patient’s tumor’s genome and compared it to a reference sequence, but also compared it to “its matched normal counterpart obtained from the same patient’s skin.” *Id.* Through this personalized genetic testing, 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. . . .” Ex. 1012 at 66. Moreover, as discussed further below, Ley discloses the method steps, and hence “[a] method for personalized genetic testing,” as claimed.

a) “sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more”

105. Ley discloses the use of whole genome sequencing to sequence nucleic acid molecules from at least one biological sample of a subject to generate sequencing data. 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 leukaemia (AML) with normal cytogenetics.” Ex. 1012 at 66.

106. Ley also discloses that the sequencing comprises sequencing 12 gigabases (GB) or more. Ley states that “[f]or the tumour genome, 32.7 ‘haploid’ coverage (98 billion bases) was obtained, and 13.9-fold coverage (41.8 billion bases) was obtained for the normal skin sample.” *Id.* Both figures—“98 billion bases,” or 98 GB, and “41.8 billion bases,” or 41.8 GB—are over the claimed 12 GB lower limit.

107. Accordingly, Ley discloses sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more.

b) “computer processing said sequencing data to identify a plurality of nucleic acid sequences having a set of genetic variants, wherein said set of genetic variants are identified with respect to a reference”

108. Ley discloses computer processing the sequencing data to identify a plurality of nucleic acid sequences having a set of genetic sequences.

109. For example, Ley reports that “[t]he analysed reads were aligned to the human reference genome (NCBI Build 36) using Maq.” *Id.* at 71 (citing Ex. 1041). One skilled in the art knew that Maq was one of many software tools available at the time to perform alignments of sequence data to a reference genome as this was

common practice in the analysis of sequence data. Indeed, the Li reference that Ley cites describes Maq as follows: “We describe the software MAQ that can build assemblies by mapping shotgun reads to a reference genome, using quality scores to derive genotype calls of the consensus sequence of a diploid, e.g., from a human sample.” Ex. 1041 at Abstract. Software like Maq, of course, is implemented on a computer and processes sequence information when it is run.

110. Ley also discloses that the sequencing revealed a large number of genetic variants with respect to a reference. In particular, a variety of “SNVs,” which are genetic variants, are described in Table 1. 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.

111. Ley also identifies a particular human reference genome: “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.” Ex. 1012 at 71.

112. Ley discloses additional computer processing of the data following an initial analysis using Maq. 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. To this end, we generated an experimental data set by re-sequencing Maq-predicted SNVs, randomly selecting a training subset and a test data set, whose annotations and features were submitted to Decision Tree C4.5 (ref. 22).” Ex. 1012 at 67-68. One skilled in the art knew that Decision Tree C4.5 was a machine learning software tool implemented on a computer and even a casual reader of Ley would understand the nature of the Decision Tree C4.5 software tool based on the title of reference 22 to which Ley cites, “Quinlan, J. R. C4.5: Programs for Machine Learning 302 (Morgan Kaufmann Publishers, 1993).” Specifically, programs for machine learning are software tools implemented on computers to aid in the analysis of large data sets.

113. Ley discloses 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. One skilled in the

art knew, as discussed above, that Maq is a sequence alignment and analysis program designed to align sequence reads to a reference and in doing so identify positions at which the sequence contains genetic variants with respect to the reference. Ley discloses that a Decision Tree is a computational algorithm that is designed to filter and identify genetic variants appropriate for precision medicine applications. In Ley's online appendix, it states:

The basic approach for the Decision Tree algorithm for SNV detection was outlined in the Results section. Specifically, SNVs were retained for additional analysis when they were supported by 7 or more unique start-site reads and where the SNV has a Maq quality score of more than 30, or if all of the following were true:

- (1) the SNV had read support with a maximum base quality greater than 26,
- (2) was supported by 3 or more unique start-site reads supporting the SNV, and
- (3) the reads supporting the SNV had an average base quality greater than 16.

Ex. 1042 at 5. In other words, Ley's Decision Tree algorithm used Maq alignments (of a certain quality score) to identify SNVs when there were 7 or more unique sequence reads and certain quality measures were satisfied.

114. Ley compares the sequencing data from the tumor and normal sample to identify ten mutations in genes that occur in the tumor but not the normal sample (referred to as "acquired" or "somatic mutations"). "A total of ten non-synonymous somatic mutations were identified in this patient's tumour genome." Ex. 1012 at 70.

c) “obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or complements thereof, wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample”

115. Ley discloses obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or complements thereof and configured to selectively enrich or amplify sequences comprising variants over other sequences. For example, Ley describes the design of PCR primers to bind to and selectively amplify the genetic variants identified in the biological sample:

By sequencing polymerase chain reaction (PCR)-generated amplicons from the tumour and skin samples (and also from the relapse tumour sample obtained 11 months after the original presentation), we determined that 152 of these variants were false positive (that is, wild type) calls, 14 were inherited SNPs, and eight were somatic mutations in both the original tumour and the relapse sample (Table 2).

Id. at 69.

116. Ley further describes the design of PCR primers to bind to and selectively amplify sequences comprising the eight somatic mutations identified in the foregoing step: “To better define the percentage of tumour cells that contained each of the discovered somatic mutations, we 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)."

Id. at 69.

117. One skilled in the art knew that sequencing PCR-generated amplicons required the design and use of nucleic acid probes designed to selectively enrich sequences comprising the set of genetic variants, such as the eight somatic variants previously described. Ley teaches that these PCR probes are designed to selectively amplify such genetic variants ("each mutation-containing locus") from the genome. *Id.* ("[W]e amplified each mutation-containing locus from non-amplified genomic DNA derived from the de novo and relapse tumour samples"). Ley further describes applying this same probe set to selectively amplify sequences in additional samples: "All of the PCR-amplified exons from the eight genes containing validated somatic mutations were sequenced in 187 further cases of AML using samples from our discovery and validation sets." *Id.* One skilled in the art knew that the coding portion of genes consist of smaller subunits known as exons and that Ley is describing amplifying and sequencing only the small portion of each gene—that exon containing the genetic variant.

d) “using said probe set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in an additional biological sample from said subject or at least one biological relative of said subject, to generate a sequencing library”

118. Ley discloses using the probe set of step (c) on an additional biological sample of the subject, *e.g.*, a sample from a “relapse tumor,” to generate a sequencing library. For example, Ley uses the eight validated 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.

e) “subjecting said sequencing library to sequencing at a coverage of 175 kilobases (kb) or less to identify a presence or absence of at least a subset of said set of genetic variants in said additional biological sample from said subject or said at least one biological relative, wherein a sequencing depth of said sequencing of (e) is greater than a sequencing depth of said sequencing of (a)”

119. Ley discloses sequencing the sequencing library:

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). Control amplicons containing a known heterozygous SNP in BRCA2 (encoding N372H) and a homozygous SNP in TP53 (encoding P72R) were analysed similarly.

Ex. 1012 at 69.

120. In Table 3, the ten genes were the panel of genes from which an exon containing a genetic variant was amplified from the original samples and from additional relapse sample from the same individual, and that these resulting amplicons were subsequently sequenced. One skilled in the art knew that the assay was designed to target only the exons of these genes containing the target genetic variants and that the use of such panels of probes was intended to amplify a small portion of the genome for subsequent sequencing rather than the genome in its entirety so that the sequencing coverage would be 175 kb or less.

121. While Ley does not provide the primer sequences or the exact exons that were analyzed, one can easily set an upper bound on the sequence coverage Ley

teaches by using the entire length of the genes that contained the genetic variant rather than the much smaller sequence of the variant-containing exon. Looking up the messenger RNA corresponding to the coding region of these ten genes at the National Center for Biotechnology Information (NCBI), one can easily obtain the length of the entire coding sequence for each of the genes. *See* Exs. 1048-59.

Gene	NCBI Accession	URL	Length (bp)
CDH24	NM_022478	https://www.ncbi.nlm.nih.gov/nuccore/NM_022478.4	3,529
SLC15A1	NM_005073	https://www.ncbi.nlm.nih.gov/nuccore/NM_005073.4	3,124
KNDC1	NM_001347864	https://www.ncbi.nlm.nih.gov/nuccore/NM_001347864.2	2,917
PTPRT	XM_047439848	https://www.ncbi.nlm.nih.gov/nuccore/XM_047439848.1	5,007
GRINL1B	AF326970.1	https://www.ncbi.nlm.nih.gov/nuccore/AF326970.1	327
GPR123	NM_001083909.3	https://www.ncbi.nlm.nih.gov/nuccore/NM_001083909.3	4,264
EBI2	NM_004951.5	https://www.ncbi.nlm.nih.gov/nuccore/NM_004951.5	1,685
PCLKC	NM_001171976.2	https://www.ncbi.nlm.nih.gov/nuccore/NM_001171976.2	4,509
FLT3	NM_004119.3	https://www.ncbi.nlm.nih.gov/nuccore/NM_004119.3	3,826
NPM1	NM_001037738.3	https://www.ncbi.nlm.nih.gov/nuccore/NM_001037738.3	1,600
BRCA2	NM_000059.4	https://www.ncbi.nlm.nih.gov/nuccore/NM_000059.4	11,954
TP53	NM_000546.6	https://www.ncbi.nlm.nih.gov/nuccore/NM_000546.6	2,512
Total coding sequence length of the ten genes in the Ley 2008 panel:			45,254

As can be seen, even if all of the exons of every one of these genes were to be sequenced, the sequence would only cover 45.254 kb, far less than the 175 kb set out as a limitation of claim 1 of the '611 patent. A person of ordinary skill in the art understood that the coding sequence of each gene represented the sum total of all the multiple exons that comprise the genes.

122. Ley further teaches sequencing the amplicons generated from the designed panel of probes to “identify a presence or absence of at least a subset of

said set of genetic variants in said additional biological sample from said subject . . .” For example, Figure 3 shows that each of the somatic mutations identified in the initial sample were also present in the relapse tumor sample:

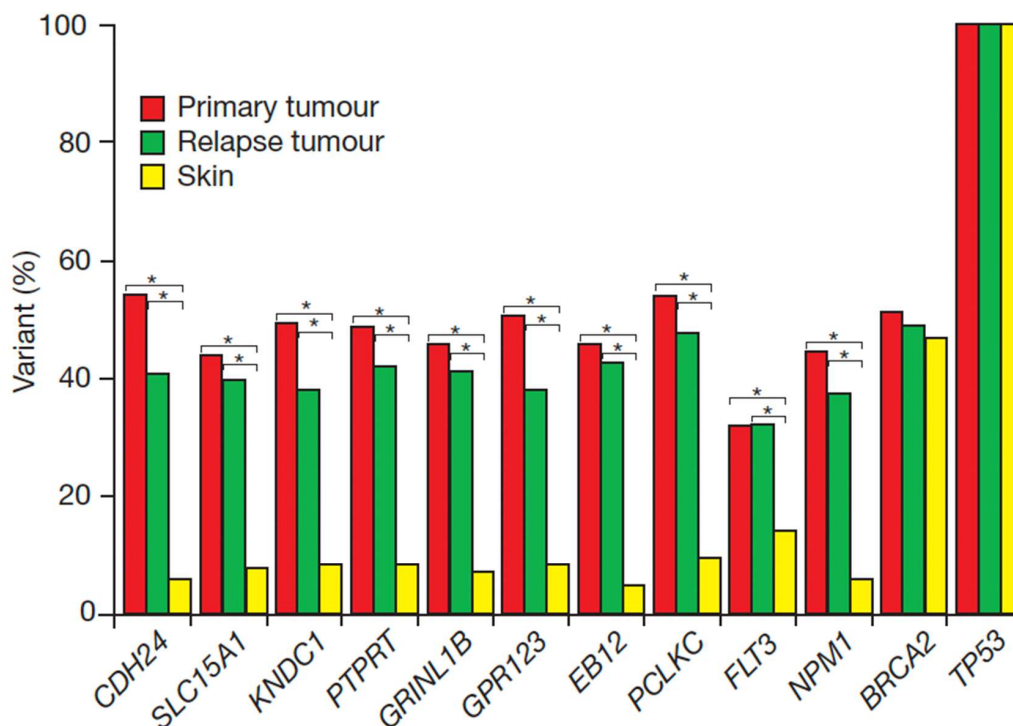


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that 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.

123. Ley also teaches the sequencing of the amplification products of the designed panel of probes, “wherein a sequencing depth of said sequencing of (e) is

greater than a sequencing depth of said sequencing of (a).” Recall that Ley teaches the sequencing of the original samples to a depth of 32.7-fold for the tumor genome and 13.9-fold for the skin genome. What these fold-coverages mean is that, on average, on average a given base was sequenced 32.7 times in the tumor and each base in skin sample was sequenced 13.9 times.

124. Table 3 lists the specific genetic variant loci that have been selectively amplified and sequenced as well as the “read counts” indicating how many times a cancer-associated genetic variant (Variant) or the “normal” reference (Ref) has been identified through sequencing of each locus.

Table 3 | 454 Readcount data for somatic mutations and known SNPs

Gene	Consequence	Primary AML (100% blasts)			Skin			Relapse (78% blasts)		
		Variant	Ref	Variant (%)	Variant	Ref	Variant (%)	Variant	Ref	Variant (%)
CDH24	Y590X	5672	4890	53.70	564	10358	5.16	3108	4599	40.33
SLC15A1	W77X	3817	4962	43.48	875	10773	7.51	4714	7173	39.66
KND1C1	L799F	4640	4848	48.90	770	8972	7.90	3883	6342	37.98
PTPRT	P1235L	998	1058	48.54	126	1489	7.80	350	493	41.52
GRINL1B	R176H	2211	2674	45.26	318	4461	6.65	1447	2070	41.14
GPR123	T38I	4618	4569	50.27	850	9751	8.02	3660	6057	37.67
EBI2	A338V	12750	15453	45.21	458	10088	4.34	2646	3627	42.18
PCLKC	P1004L	992	855	53.71	341	3153	9.76	705	773	47.70
FLT3	ITD	4220	7810	35.08	3475	23159	13.05	3870	8495	31.30
NPM1	CATG ins	1550	1974	43.98	143	2390	5.65	2303	3910	37.07
BRCA2	N372H	778	752	50.85	763	876	46.55	285	303	48.47
TP53	P72R	8989	1	99.99	8161	0	100.00	7914	6	99.92

The differences between variant frequencies in primary or relapse tumour samples and skin were highly significant for all somatic mutations ($P < 0.000001$, Fisher's exact test, one tailed). The BRCA2 variant is a known heterozygous SNP in this genome, and the TP53 variant is a known homozygous SNP.

125. Summing the variant and reference read counts for each gene in each sample, one can see that the variant loci in the additional relapse sample (“additional biological sample from said subject”) have been sequenced at depths ranging from 588-fold to 12,365-fold – far greater than the 37.2-fold coverage of the original samples.

Gene	Primary AML			Skin			Relapse		
	Variant	Ref	Total Depth	Variant	Ref	Total Depth	Variant	Ref	Total Depth
CDH24	5672	4980	10652	564	10358	10922	3108	4599	7707
SLC15A1	3817	4962	8779	875	10773	11648	4714	7173	11887
KNDC1	4640	4848	9488	770	8972	9742	3883	6342	10225
PTPRT	998	1058	2056	126	1489	1615	350	493	843
GRINL1B	2211	2674	4885	318	4461	4779	1447	2070	3517
GPR123	4618	4569	9187	850	9715	10565	3660	6057	9717
xxEBI2	12750	15443	28193	458	10088	10546	2646	3627	6273
PCLKC	992	855	1847	341	3153	3494	705	773	1478
FLT3	4220	7810	12030	3475	23159	26634	3870	8495	12365
NPM1	1550	1974	3524	143	2390	2533	2303	3910	6213
BRCA2	778	752	1530	763	876	1639	285	303	588
TP53	8989	1	8990	8161	0	8161	7914	6	7920

2. Claim 2

126. Claim 2 recites “[t]he method of claim 1, further comprising outputting a report that is generated at least based on comparison of results from said sequencing of (a) with results from at least said subjecting said sequencing library to said sequencing of (e).”

127. Ley’s “Methods Summary” explains the following workflow:

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. Coverage of the tumour and normal genomes was ascertained by comparison to the patient’s heterozygous SNPs, established by compiling shared SNP calls monitored on the Affymetrix 6.0 and Illumina Infinium 550K

geno- typing platforms. We 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. ***For all putative variants, we attempted validation using custom PCR and capillary sequencing on the ABI 3730 platform. All validated somatic mutations were further analysed by Roche/ 454 sequencing of PCR-generated amplicons made from primary genomic DNA to compare readcounts of wild-type and mutant alleles in the primary tumour, skin and relapse tumour samples.***

Ex. 1012 (Ley) at 71.

128. The first part of this workflow involves whole-genome sequencing with the Illumina Genome Analyzer and Solexa sequencing platform to identify variants of interest. Specifically, Ley describes “the use of whole-genome re-sequencing and variant discovery approaches using the Illumina/Solexa technology . . . to sequence a tumor genome of a single AML patient and the matched normal genome (derived from a skin biopsy) of the same patient.” Ex. 1012 (Ley) at 67. Ley describes the details of this whole-genome sequencing step as follows:

Because most of the acquired mutations in cancer genomes have been shown to be heterozygous, the complete sequencing of a cancer genome requires the detection of both alleles at most positions in the genome. We therefore designed sequence coverage metrics to define the point at which 90% diploid coverage had been reached. To minimize errors associated with any single platform or measurement, diploid coverage

for this genome was assessed using a set of high- quality SNPs derived from two different SNP array platforms, Affymetrix 6.0 and Illumina Infinium 550K. For a SNP to be included in the high-quality set, the following criteria had to be satisfied: (1) identical genotypes were called from both assays at the same genomic positions, and (2) the resulting genotype was heterozygous. For the 933124 tumour genome, 46,494 heterozygous SNs passed the above criteria and were defined as high-quality SNPs. For the skin samples, 46,572 high-quality SNPs were defined.

We performed 98 full runs on the Illumina Genome Analyser to achieve the targeted level of 90% diploid coverage as determined by coverage of the high-quality SNP set. . . .

We sequenced the genome of normal skin cells from the same patient to enable the identification of inherited sequence variants in the tumour genome. Our targeted diploid coverage goal for the skin-derived genome was 80%. We achieved this goal with only 34 Solexa runs (41.8 billion bases), using improved reagents and longer read lengths to attain 82.6% diploid and 84.2% haploid coverage (Table 1).

Ex. 1012 (Ley) at 67 (“Coverage depth of the tumour and skin genomes”).

129. Ley further discloses generating a report, for example, through tabular summary of results. Specifically, Table 1, as referenced in the excerpt above, is a table summarizing various results of sequencing 2 including reported information at least based on comparison of results from said sequencing of (a) (e.g., “Runs” of

“Tumour” and “Skin” samples) with results from at least said subjecting said sequencing library to said sequencing of (e) (e.g.. “SNVs detected with respect to hg18 (no Y)” for “Tumor” and “Skin” samples). *Id.* at 68.

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.

130. Ley’s Figure 3 also provides a report depicting 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.

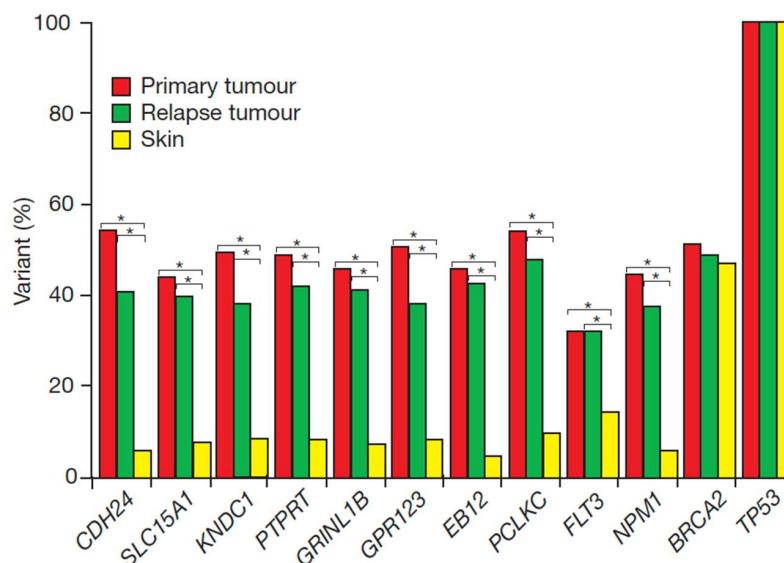


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that 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.

3. Claim 4

131. Claim 4 recites “[t]he method of claim 1, wherein said sequencing of (a) comprises (i) exome sequencing, (ii) sequencing a panel of genes, (iii) whole genome sequencing, and/or (iv) sequencing a population of complementary deoxyribonucleic acid molecules derived from ribonucleic acid molecules.”

132. As explained above with respect to claim 2, Ley's “sequence data of (a)” is sequence information from whole genome sequencing. As Ley states, “[o]ur 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.* at 66.

4. Claim 5

133. Claim 5 recites “[t]he method of claim 1, wherein each of said plurality of nucleic acid probe molecules includes oligonucleotide-directed genomic content comprising (i) at least one variable portion from a result of said sequencing of (a) and (ii) at least one fixed portion independent of said result of said sequencing of (a).”

134. Ley’s personalized probe set comprises nucleic acid probe molecules—specifically PCR primers—that are directed to the genes described on the personalized tumor panel. *See* Ex. 1012 (Ley) at 71 (“For all putative variants, we attempted validation using custom PCR and capillary sequencing on the ABI 3730 platform. All validated somatic mutations were further analyzed by Roche/454 sequencing of PCR-generated amplicons made from primary genomic DNA to compare readcounts of wild-type and mutant alleles in the primary tumour, skin and relapse tumour samples.”). Such PCR probes, by their nature, contain oligonucleotide-directed genomic content, specifically oligonucleotide sequences that base-pair or hybridize to their complementary sequences. *See* Section V.C.3.a).

135. Ley’s PCR primers comprise the claimed “at least one variable portion from a result of said sequencing.” Specifically, Ley discloses including primers for

the non-synonymous somatic mutations detected in the AML sample. This part of Ley's panel is "variable . . . from a result of said sequence data" at least because "eight [of ten] somatic mutations that [Ley] detected are all single base changes, **and none has previously been detected in an AML genome.**" Ex. 1012 (Ley) at 70; *see also id.* at 69 ("To better define the percentage of tumour cells that contained each of the discovered somatic mutations, we 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."); *id.* (Table 2); *id.* at 71 ("We cannot overemphasize the importance of parallel sequencing of the patient's normal genome to determine which variants were inherited; the identification of the true somatic mutations in this tumour genome would not have been possible without this approach.").

136. Ley's PCR primers also comprises "at least one fixed portion independent of said result of said sequencing of (a)." Specifically, Ley discloses including primers for known mutations of the BRCA2 and TP53 genes. "The BRCA2 variant is a known heterozygous SNP in this genome, and the TP53 variant is a known homozygous SNP." *Id.* at 70 (Table 3, caption); *see also id.* at 69 ("Control amplicons containing a **known** heterozygous SNP in *BRCA2* (encoding

N372H) and a homozygous SNP in *TP53* (encoding P72R) were analyzed similarly.”). The literature also confirmed that *BRCA2* and *TP53* were known AML tumor variants. *See* Ex. 1073 (Haferlach 2008) and Ex. 1074 (Misawa 1998) (confirming knowledge of such variants). The inclusion of these control amplicons were based on known genetic variants, which result in known changes in the protein encoded, at a known gene dosage (i.e., heterozygous – resulting in half of the protein produced having that mutation, or homozygous, resulting in all of the protein produced by the mutated gene having that mutation). The identity and effect of these mutations was known to the authors prior to the whole-genome sequencing step, reflected by the selection of these variants because they contained “known” SNPs that occur in all of the patient’s cells (i.e., germline variants). Thus, these control amplicons constitute “fixed” primers independent of the variable primers that are selected and designed based on the identity of the SNPs and other mutations identified via whole-genome sequencing.

5. Claim 9

137. Claim 9 recites “[t]he method of claim 1, further comprising outputting a report that identifies said presence or absence of said at least said subset of said genetic variants in said subject or said at least one biological relative.”

138. For the same reasons discussed above for claim 2, Ley discloses the limitations of claim 9.

6. Claim 10

139. Claim 10 recites “[t]he method of claim 1, further comprising providing a therapeutic intervention at least based on said presence or absence of said at least said subset of said genetic variants identified in (e).”

140. In addition to the reasons given above for claim 2, Ley discloses claim 10. Ley’s report discloses genetic variants which inform a therapeutic intervention. Specifically, “[a]lthough the patient experienced a complete remission with conventional therapies, she relapsed at 11 months and expired 24 months after her initial diagnosis was made. At relapse the bone marrow had 78% myeloblasts, and contained a new clonal cytogenetic abnormality, t(10;12)(p12;p13).” Ex. 1012 (Ley) at 67. Thus, Ley’s report, which shows elevated levels of certain variants, informs a therapy choice or a change (*e.g.*, switching away from conventional therapy, or resuming therapy upon relapse) for the subject. For example, in the Supplementary Information, Ley presents the clinical and therapeutic implication of the SLC15A1. “SLC15A1 is a transmembrane peptide transporter that is highly related to SLC11A1 (NRAMP1), SLC19A1, and SLC29A1 (ENT1), all of which are implicated in the transport of small molecules and drugs to cancer cells. Biallelic mutations in SCL29A1 that result in a lack of detectable protein recently have been associated with cytarabine resistance in a human leukemia cell line, implicating SLC family members in chemotherapy resistance for AML.” Ex. 1042 at 9. One skilled in the art

would recognize that cytarabine is a chemotherapy drug use in the treatment of AML and that the presence of the SCL29A1 would inform an oncologist that the use of this drug was contraindicated.

141. Further, one skilled in the art knew that the characteristics of the disease and the patient's positive mutational in the FLT3 and NMP1 would contraindicate the use of allogenic stem cell transplant for this patient. "Based on available data, all adults with AML under age 60 years with matched siblings should be considered for allogeneic transplantation in first remission, except for those with favorable risk cytogenetics and possibly those whose disease has normal cytogenetics and is FLT3/ITD negative and NPM1 positive." Ex. 1075 at 162 (Davies 2007).

142. Moreover, as explained above in Section V.F, an important aspect of reporting the results of cancer-monitoring assays is providing actionable information to healthcare providers. Thus, a POSA knew how to generate "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." Ex. 1014 (Ross) at 537. Ley's Figure 3 is an example of such a result report which shows the presence or absence of a health condition at different time points over the course of the patient's cancer.

7. Claim 11

143. Claim 11 recites “[t]he method of claim 1, wherein said at least one biological sample includes a tumor sample and said nucleic acids molecules are from cells in said tumor sample.”

144. As discussed above, Ley’s initial whole-genome sequencing data is derived from sequences of nucleic acids taken from a primary tumor sample, *i.e.*, a tumor biopsy.

8. Claim 13

145. Claim 13 recites “[t]he method of claim 1, wherein said plurality of nucleic acid sequences having said set of genetic variants comprises genetic variants in a germline sequence of said subject.”

146. Ley discloses this limitation, for example, as follows: “We sequenced the genome of normal skin cells from the same patient to enable the identification of inherited sequence variants in the tumour genome.” Ex. 1012 at 67. Normal skin cells have germline sequence as recited in the claim. For example, as shown in Ley’s Figure 3, the subject’s germline sequence had genetic variants in the BRCA2 and TP53 genes. *Id.* at 69.

9. Claim 18

147. Claim 18 recites “[t]he method of claim 1, wherein subjecting said sequencing library to sequencing in (e) comprises generating nucleic acid sequence data from said additional biological sample.”

148. The sequencing library was created by amplifying the relapsed tumour sample with the set of PCR primers and the resulting amplicons were then sequenced. Ex. 1012 (Ley) at 69 (“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).”).

10. Claim 19

149. Claim 19 recites “[t]he method of claim 1, wherein computer processing of said sequence data comprises identifying one or more of (i) genetic variants of said nucleic acid sequence with respect to a germline sequence(s), (ii) alleles which match said reference sequence(s) and are correlated with a type of cancer or other disease, (iii) alleles which determine a human leukocyte antigen (HLA) type, (iv) metrics of gene expression or allele-specific expression, and (v) quantification of non-coding ribonucleic acid (RNA) molecules or micro-RNA molecules which are at least partially tissue-type specific or cancer-type specific.”

150. As discussed above, the *BRCA2* and *TP53* genes included in Ley’s fixed panel are alleles which match said reference sequence(s) and are correlated

with a type of cancer or other disease. Further, as discussed above, Ley's computer processing-based comparison of the whole genome sequence data from the primary tumor to the patient's germline sequence identified a number of somatic mutations (acquired "genetic variants" found in the primary tumor) "with respect to a germline sequence." *See* Ex. 1012 at 69, Figure 3.

11. Claim 28

151. Claim 28 recites "[t]he method of claim 1, further comprising using said presence or absence of said at least said subset of said set of genetic variants identified in (e) to treat said subject or said at least one biological relative for a disorder."

152. For the reasons given above for claim 10, Ley discloses the limitations of claim 28. Ley's report also discloses genetic variants which inform treating a subject. Specifically, "[a]lthough the patient experienced a complete remission with conventional therapies, she relapsed at 11 months and expired 24 months after her initial diagnosis was made. At relapse, the bone marrow had 78% myeloblasts, and contained a new clonal cytogenetic abnormality, t(10;12) (p12;p13)." Ex. 1012 (Ley) at 67. Thus, Ley's report (*e.g.*, Figure 3), which shows elevated levels of certain variants, informs a therapy choice or a change (*e.g.*, switching away from conventional therapy, or resuming therapy upon relapse) for the subject.

153. Ley's Supplementary Information, Ex. 1042, provides a detailed case report and provides detailed information about a number of the various mutations that had been detected in the patients cancer genome. The germline and somatic genetic variants detected in the patient are linked by Ley to the diagnosis of AML. "Germline copy number variants (CNVs) also may have contributed to AML susceptibility in this patient. Among the genes that lie within intervals constitutionally deleted in this individual are WWOX, a tumor suppressor somatically deleted or translocated in a variety of solid tumors."

154. In the Supplementary Information, Ley also presents the clinical and therapeutic implication of the SLC15A1. "SLC15A1 is a transmembrane peptide transporter that is highly related to SLC11A1 (NRAMP1), SLC19A1, and SLC29A1 (ENT1), all of which are implicated in the transport of small molecules and drugs to cancer cells. Biallelic mutations in SCL29A1 that result in a lack of detectable protein recently have been associated with cytarabine resistance in a human leukemia cell line, implicating SLC family members in chemotherapy resistance for AML." Ex. 1042 at 9. One skilled in the art would recognize that cytarabine is a chemotherapy drug use in the treatment of AML and that the presence of the SCL29A1 would inform an oncologist that the use of this drug was contraindicated.

155. Further, one skilled in the art knew that the characteristics of the disease and the patient's positive mutational in the FLT3 and NMP1 would contraindicate

the use of allogenic stem cell transplant for this patient. "Based on available data, all adults with AML under age 60 years with matched siblings should be considered for allogeneic transplantation in first remission, except for those with favorable risk cytogenetics and possibly those whose disease has normal cytogenetics and is FLT3/ITD negative and NPM1 positive." Ex. 1075 at 162 (Davies 2007).

156. [Intentionally blank paragraph]

12. Claim 29

157. Claim 29 recites "[t]he method of claim 1, wherein said reference comprises a human reference sequence."

158. Ley discloses this limitation, for example, as follows: "To begin evaluating the quantity and quality of the detected sequence variants in the tumour and skin genomes, we compared the overlap and uniqueness of this genome's variants with respect to the James D. Watson and J. Craig Venter genomes, and to dbSNP (v127; Fig. 1)." Ex. 1012 at 67.

13. Claim 30

159. Claim 30 recites "[t]he method of claim 1, wherein said reference is a human reference sequence."

160. Ley discloses this limitation for the same reasons as claim 29.

B. Ground 2: Ley Renders Obvious Claims 1-5, 9-11, 13, 18, 19, 28-30

161. Claims 1-5, 9-11, 13, 18, 19, 28-30 would have been obvious over Ley (Ex. 1012), as discussed for Ground 1 and further discussed below for Ground 2.

162. Claim 1 recites:

A method for personalized genetic testing, comprising:

(a) sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more;

(b) computer processing said sequencing data to identify a plurality of nucleic acid sequences having a set of genetic variants, wherein said set of genetic variants are identified with respect to a reference;

(c) obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or complements thereof, wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample;

(d) using said probe set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in an additional biological sample from said subject or at least one biological relative of said subject, to generate a sequencing library; and

(e) subjecting said sequencing library to sequencing at a coverage of 175 kilobases (kb) or less to identify a presence or absence of at least

a subset of said set of genetic variants in said additional biological sample from said subject or said at least one biological relative, wherein a sequencing depth of said sequencing of (e) is greater than a sequencing depth of said sequencing of (a).

1. Claims 1-5, 9-11, 13, 18, 19, 28-30

163. To the extent claims 1-5, 9-11, 13, 18, 19, or 28-30 are not anticipated as explained in Ground 1, they would have been obvious in view of Ley.

164. To the extent Ley's disclosure not anticipate the limitation in step 1(e) concerning "sequencing at a coverage of 175 kilobases (kb) or less," a person of ordinary skill would have found it obvious to sequence at a coverage of 175 kilobases (kb) or less based on Ley's disclosure.

165. As a preliminary matter, the '611 patent does not attach any significance to the number "175 kilobases." Instead, the patent teaches that 175 kilobases is simply far less than the about 35 million bases that are typically sequenced in exome sequencing. Ex. 1001 at 35:1-5. But 150 kilobases or 200 kilobases or any other number significantly less than 35 million bases are also far less than about 35 million bases.

166. If a person of ordinary skill would not have recognized express disclosure in Ley of coverage of 175 kilobases or less, then modifying Ley's

disclosure to provide coverage (a capture panel) of 175 kilobases or less would have been obvious for that reason.

167. A person of ordinary skill knew that achieving a higher sequence depth in step 1(e) was desirable for increasing accuracy of detecting and quantifying mutations in tumors. For a given number of sequencing counts, the use of fewer oligonucleotide probes to target smaller genomic regions would lead to higher sequencing depth, as would be expected from the Lander/Waterman equation. In addition, even though sequencing had become less expensive over the years prior to the '611 patent, a person of ordinary skill knew reducing sequencing costs further was desirable, and that smaller sequencing panels in step 1(e) could reduce sequencing costs.

168. A person of ordinary skill would have reasonably expected to be successful in reducing coverage to 175 kb or less by choosing an informative subset of genes to assay. Providing capture panels of smaller coverage was easier, in fact, than greater coverage, because smaller coverage was achieved by having fewer oligonucleotide probes—a simplifying factor in capture panel design and production. A smaller panel also simplified computational analysis because the resulting data would be limited to the few genes that were represented in a personalized panel (of the approximately 25,000 genes encoding in the genome).

2. Claim 3—Further Information Regarding Obviousness

169. Claim 3 recites “[t]he method of claim 1, wherein in said sequencing of (a), said at least one biological sample is obtained from said subject at a first time point, and wherein said additional biological sample is obtained from said subject or said at least one biological relative at a second time point subsequent to said first time point.”

170. Given that a benefit of the assay is the ability to observe the course of a patient’s cancer and response to treatment, a POSA would have found it obvious to repeat the assay at a second time point, resulting in the collection of a second biological sample following the initial sample for analysis and sequencing using the Ley’s personalized probe sets. A POSA would have reasonably expected success in analyzing multiple samples, given that it would have involved merely applying the techniques already proven to be successful at multiple time points. Indeed, soon after, Leary (Ex. 1002) and McBride (Ex. 1015) demonstrated the use of a similar assay design, by using whole-genome sequencing to identify genetic 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 plasma of patients with solid tumors at multiple (e.g., 6) time points. *See also* Ex. 1031 (Meyerson) at 687 (Table 1) (listing Ley and subsequent whole cancer genome studies).

3. Claim 5—Further Information Regarding Obviousness

171. Claim 5 recites “[t]he method of claim 1, wherein each of said plurality of nucleic acid probe molecules includes oligonucleotide-directed genomic content comprising (i) at least one variable portion from a result of said sequencing of (a) and (ii) at least one fixed portion independent of said result of said sequencing of (a).”

172. In addition to the PCR probes comprising the fixed portion already included in Ley’s panel, a POSA reading Ley in view of the knowledge in the field would have also found it obvious to include probes targeted to genetic variants involved in the pharmacogenomics of cancer drugs, *i.e.*, genetic variants that were known to have companion diagnostics, as well as to tumor-driver mutations that were known in the art.

173. For example, as explained in Section V.F (“Providing Actionable Information to Healthcare Providers”) above, many effective cancer therapies are targeted to particular genetic mutations that present in certain forms of cancer. Examples include the blockbuster drug Herceptin targeting HER2 positive breast cancer (*i.e.*, forms of breast cancer carrying mutations in the HER2 gene), which was first introduced in 1998. *See* Ex. 1014 (Ross 2011) at 527 (“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.”).

174. Relatedly, the identity of therapy/cancer pairs would have been readily ascertainable by one skilled in the art by accessing the treatment guidelines for a particular cancer promulgated by the National Comprehensive Cancer Network (called the NCCN Guidelines) which are widely used by oncologists to reference and define standard of care treatments. Similar information would have also been available from the FDA-approved prescribing information available for relevant drugs. To design a probe for relevant mutations identified in this manner, a POSA would have been able to determine the relevant sequence information from any one of the publicly available cancer databases, such as COSMIC, *see* Section V.B (“Genetic Mutations”), and then designed nucleic acid probes based on the techniques already known in the art and applied in the methods expressly disclosed by Ley. Accordingly, a POSA would have reasonably expected success in the predictable use of known information (regarding the sequences) and known techniques (the creation of nucleic acid probes to target such sequences) to add

“fixed” probes to the types of assays described in Ley in order to provide information relevant to the selection and response to mutation-specific therapy pairings.

175. Additionally, as discussed above, a POSA would have been familiar with known mutations common in certain forms of cancer, which are recorded in databases based on population-level studies. Applying PCR-based or capture-based targeting approaches to such mutations using nucleic acid probe molecules was known and discussed by Ley: “[H]ypothesis-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.” Ex. 1012 (Ley) at 70. A POSA would have been motivated to combine candidate gene-based probes (fixed portion) in a panel with patient specific mutations discussed above (variable portion) in order to assess the cancer for known mutations in well-characterized and studied candidate genes in addition to patient-specific mutations. For example, as noted above, in many instances such genes may drive the course of the cancer and may be key targets for therapy. Further, such a combination of fixed and variable probes would overcome the limitations identified by Ley of a purely hypothesis-driven (fixed probe-only) approach.

176. Ley’s observation that the use of such probes may be limited does not teach away from including them in a tumor-monitoring panel. In fact, based on the knowledge of cancer biology gained between Ley’s publication and the priority date,

a POSA would have known that one reason a targeted approach may miss such mutations is because of tumor heterogeneity. Tumor heterogeneity refers to the fact that genetic variants present in a cancer are not always uniformly dispersed across a tumor spatially, and that certain key variants may not be present in the tumor at all times during the cancer's progression. *See, e.g.,* Ex. 1071 (Burrell) at 338 (“Extensive genetic and phenotypic variations exist not only between tumours (intertumour heterogeneity) but also within individual tumours (intratumour heterogeneity; Fig. 1). In tumours, populations of subclones can intermingle or be spatially separated, and this subclonal architecture varies dynamically throughout the disease course.”). This explains why sequencing a tumor sample may miss some of the patient's tumor-associated mutations at any given point in time along the course of the disease. As a result, it would have been obvious to include probes for selectively enriching such hotspot mutations (*i.e.*, the claimed fixed probes) in the panel in order to detect in subsequent samples the mutations previously identified via the WGS assay, but also mutations that were known to be common in the patient's type of cancer. A POSA would have been motivated to monitor hotspot mutations for a particular cancer (*i.e.*, the fixed probes) even in cases where they may not have been identified at the time of initial presentation or at the time of an initial whole-genome sequencing assay, as such mutations may present later in the course of the patient's cancer.

177. A POSA would have been able to readily create probes for such mutations using already known techniques. Indeed, Ley already recognized that such probes were well-known in the art, and thus would have reasonably expected success in including such “fixed” probes in a cancer-monitoring panel. Further, such fixed probes would selectively amplify sequences comprising genetic variants in exactly the same way as Ley’s variable probes selectively amplified sequences comprising patient-specific genetic variants.

4. Claim 28—Further Information Concerning Obviousness

178. Claim 28 recites “[t]he method of claim 1, further comprising using said presence or absence of said at least said subset of said set of genetic variants identified in (e) to treat said subject or said at least one biological relative for a disorder.”

179. Claim 28 is obvious over Ley. As explained above in Section V.F, an important aspect of reporting the results of cancer-monitoring assays is providing actionable information to healthcare providers. A POSA would have been motivated to generate “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.” Ex. 1014 (Ross) at 537. Ley’s Figure 3 is an example of such a result report, which shows the presence or absence of a health condition at different time points over the course of the

patient's cancer. It would have been obvious to report similar information at each time point for which a sample was collected and analyzed in Ley's assays as well as in assays that were obvious variants of those disclosed in Ley. Moreover, it would have been obvious to identify genetic variants which inform a therapy choice or change in therapy in such a report, in particular in instances in which the panel of genetic variants included probes designed to detect and quantify genes associated with a companion therapeutic, as discussed above

Ground 3: Newman Anticipates Claims 1-5, 9-11, 13, 18, 19, 28-30

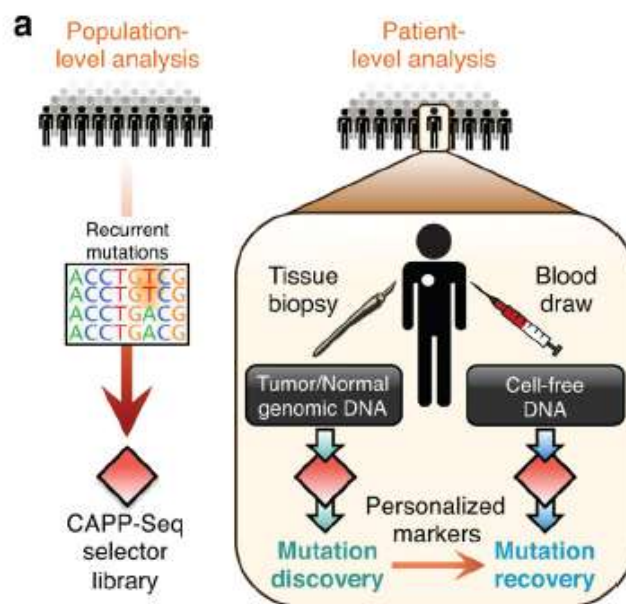
180. Newman discloses a method for an improved “cancer personalized profiling by deep sequencing (CAPP-Seq).” Ex. 1043 at 547. Newman's CAPP-Seq method, which “combines hybrid affinity capture of hundreds of genomic regions with deep sequencing and a specialized bioinformatics workflow.” *Id.* Newman reports that its method “substantially outperforms ... [other] methods for the analysis of ctDNA from clinically practical blood volumes and allows for efficient noninvasive ctDNA profiling in the absence of available tissue biopsies.” *Id.* Newman's approach includes the following steps.

181. ***First***, the laboratories that published Newman previously published an article disclosing the CAPP-Seq process. Ex. 1044. In that prior article, the Newman laboratories taught that

CAPP-Seq, combines optimized library preparation methods for low DNA input masses with a multi-phase bioinformatics approach to design a “selector” consisting of biotinylated DNA oligonucleotides that target recurrently mutated regions in the cancer of interest. To monitor ctDNA, the selector is applied to tumor DNA to identify a patient’s cancer-specific genetic aberrations and then directly to circulating DNA to quantify them (Fig. 1a).

Ex. 1044 (Newman 2014) at 2 (“Introduction”).

Figure 1 depicts the design of the “selector” and its use for ctDNA:



182. **Second**, Newman applied its method to samples from a glioblastoma (GBM) tumor using a 157-kb personalized selector followed by sequencing to cover mutations in the tumor that Newman discovered in its patient-level analysis of the

tumor tissue. More specifically, Newman performed the patient-level analysis with whole exome sequencing of the glioblastoma tumor. Ex. 1043 at “ctDNA detection limits for iDES, duplex sequencing, and other methods” in “Online Methods.” That sequencing information was used to create the 157-kb personalized selector including hybrid capture baits that selectively enriched the target sequences of the GBM tumor. Newman “then applied this selector to perform deep resequencing of the GBM tumor (1/2 of a MiSeq lane), and confirmed 97% of expected SNVs, 99.7% of which had duplex support (n = 1502).” *Id.*

183. As explained in more detail below, Newman anticipates Claims 1-5, 9-11, 13, 18, 19, and 28-30 of the '611 patent.

5. Claim 1

184. Claim 1 recites:

A method for personalized genetic testing, comprising:

(a) sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more;

(b) computer processing said sequencing data to identify a plurality of nucleic acid sequences having a set of genetic variants, wherein said set of genetic variants are identified with respect to a reference;

(c) obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or

complements thereof, wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample;

(d) using said probe set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in an additional biological sample from said subject or at least one biological relative of said subject, to generate a sequencing library; and

(e) subjecting said sequencing library to sequencing at a coverage of 175 kilobases (kb) or less to identify a presence or absence of at least a subset of said set of genetic variants in said additional biological sample from said subject or said at least one biological relative, wherein a sequencing depth of said sequencing of (e) is greater than a sequencing depth of said sequencing of (a).

185. The preamble of claim 1 recites “[a] method for personalized genetic testing.” Newman discloses it by pointing out that “[h]igh-throughput sequencing of circulating tumor DNA (ctDNA) promises to facilitate personalized cancer therapy.” Ex. 1043 at 1 (Abstract).

186. Moreover, as discussed further below, Newman discloses the method steps, and hence “[a] method for personalized genetic testing,” as claimed.

a) “sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more”

187. Newman discloses sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data. For example, Newman terms its personalized sequencing approach CAPP-Seq, or cancer personalized profiling by deep sequencing. Ex. 1043 at 1 (Abstract). As discussed above, the same laboratories previously published the development of the CAPP-Seq methods. Ex. 1044 (Newman 2014).

188. Newman discloses that the sequencing comprises sequencing 12 gigabases (GB) or more. As an initial matter, the '611 patent does not ascribe any inventiveness to sequencing 12 gigabases or more. The '611 patent discusses 12 gigabases only once in an example—Example 2—by stating that “a germline DNA sample can be exome sequenced (e.g., 12 billion bases).” Ex. 1001 at 62:33-35. The use of “e.g.” would have been understood by a person of ordinary skill to mean that 12 billion bases (gigabases) was simply an example of how much exome sequencing could be used for the '611 patent method. As noted above, the range for whole exome sequencing was about 0.2 GB to 20 GB. Ex. 1040 at 123, Figure 1.

189. Newman discloses that “[s]equencing was performed using 2 x 100 or 2 x 150 paired-end reads” Ex. 1043 at “Library preparation and sequencing” in

“Online Methods.” In its supplementary tables, published online together with and at the same time as the primary Newman article, Newman discloses data from whole exome sequencing—the same type of sequencing as in the ’611 patent Example 2 discussing 12 GB of sequencing—of two samples from a glioblastoma patient. Ex. 1045 at Table 2, tab entitled “SuppTable2_samples_and_QC”, rows 286-87. In particular Newman reports that these two whole exome sequencing runs resulted in totals of 82,247,848 and 86,781,192 reads respectively. *Id.*

Sample number	Cohort	Case	Collection time ^a	Sample type	Input mass (ng)	Selector ^b	Sequencing adapter ^c	De-duplication ^d	Total reads
275	Healthy/Te	GBM-CTR25	GBM-CTR25	cfDNA	32.0	GBM	Tandem	Barcoding	37,647,656
276	Healthy/Te	GBM-CTR26	GBM-CTR26	cfDNA	32.0	GBM	Tandem	Barcoding	38,321,932
277	Healthy/Te	GBM-CTR27	GBM-CTR27	cfDNA	32.0	GBM	Tandem	Barcoding	33,065,346
278	Healthy/Te	GBM-CTR28	GBM-CTR28	cfDNA	32.0	GBM	Tandem	Barcoding	35,976,502
279	Technical	GBM-Spike-0	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	103,288,518
280	Technical	GBM-Spike-0	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	105,786,270
281	Technical	GBM-Spike-0	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	100,222,542
282	Technical	GBM-Spike-0	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	66,777,970
283	Technical	GBM-Spike-0	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	56,060,804
284	Technical	HD500-0.5%-i	HD500-0.5%-i	Cell line bli	72.0	NSCLC tumor ge	Tandem	Barcoding	270,879,108
285	Healthy/Te	CTR31	CTR31-c1	cfDNA	32.0	NSCLC tumor ge	Tandem	Barcoding	259,725,374
286	GBM	GBM Relapse	na	Tumor	100.00	SeqCap EZ Hum.	Tandem	Barcoding	82,247,848
287	GBM	GBM PBL	na	PBL	100.00	SeqCap EZ Hum.	Tandem	Barcoding	86,781,192

Based on the 150 bp read length, the sequencing comprises sequencing 12.3 GB (150 * 82,247,848) and 13.0 GB (150 * 86,781,192) respectively. Even based on the shorter 100 bp read length, the claim language instructs that one may combine the two samples (“sequencing nucleic acid molecules from *at least one biological sample* of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more”), which would result in sequencing comprising (8.2 GB (100 * 82,247,848)) + (8.7 GB (100 * 86,781,192)) = 16.9 GB.

190. Accordingly, Newman discloses sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more.

b) “computer processing said sequencing data to identify a plurality of nucleic acid sequences having a set of genetic variants, wherein said set of genetic variants are identified with respect to a reference”

191. Newman discloses computer processing the sequencing data to identify a plurality of nucleic acid sequences having a set of genetic sequences. For example, for the glioblastoma tumor that Newman studied, Newman performed whole exome sequencing and found it to “harbor ~1,500 coding mutations.” Ex. 1043 at “ctDNA detection limits for iDES, duplex sequencing, and other methods” in “Online Methods.” These mutations are genetic variants in the tumor, and a person of ordinary skill would have understood Newman to be disclosing the use of computer processing to identify them because such large sequencing data sets were (and still are) always analyzed using computer processing in the field. *See* Ground 1, claim 1(b).

192. Also, the variants were identified with respect to a reference, *i.e.*, the unmutated sequences corresponding to the locations of the ~1,500 coding mutations. Newman’s CAPP-Seq process, which was used for all the testing including for the glioblastoma tumor, includes producing a selector containing capture-based bait

oligonucleotides corresponding to polymorphisms in a tumor. The selection of oligonucleotide probe sequences requires comparison to a reference sequence (*i.e.*, the germline sequence for the GBM patient).

c) **“obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or complements thereof, wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample”**

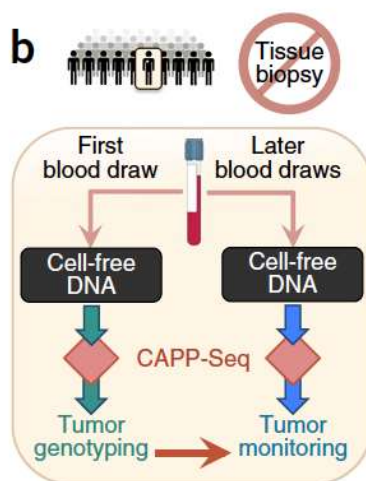
193. Newman discloses obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or complements thereof. For example, Newman states that “[w]e designed a 157-kb ‘**personalized**’ selector to cover these [~1,500] mutations, employing the same biotinylated oligonucleotide synthesis process as described for the NSCLC selector design, above.” Ex. 1043 at “ctDNA detection limits for iDES, duplex sequencing, and other methods” in “Online Methods.” For the referenced NSCLC selector design, Newman states that “[b]iotinylated oligonucleotide capture pools for both selectors were designed using the NimbleDesign portal (Roche NimbleGen) and genome build hg19 (NCBI Build 37.1/GRCh37).” *Id.* at “NSCLC selector design” in “Online Methods.”

194. Newman’s “[b]iotinylated oligonucleotide capture pools ... designed using the NimbleDesign portal (Roche NimbleGen)” are configured to selectively

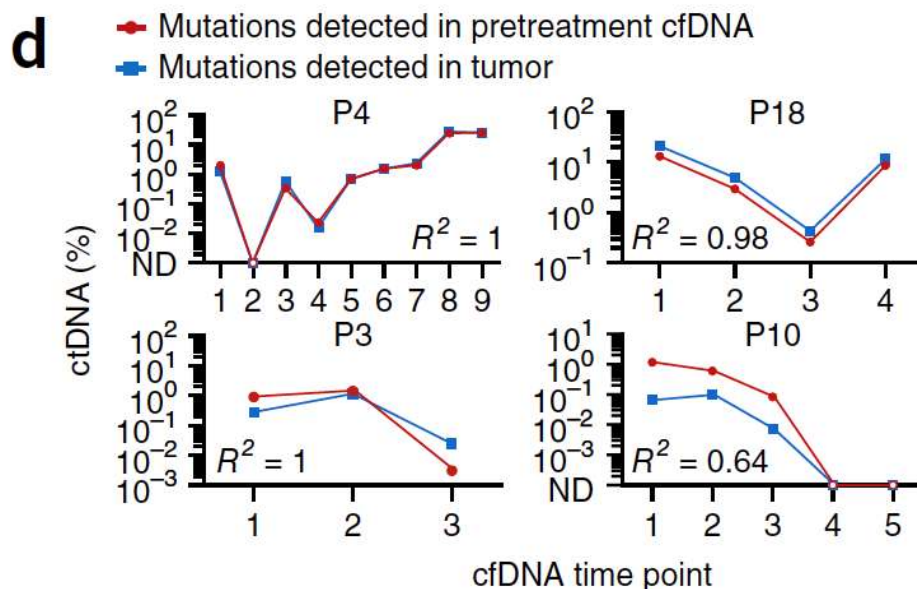
enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample. For example, the NimbleGen system was known to be a commercially-available exome enrichment platform that used “biotinylated oligonucleotide baits complementary to the exome targets” Ex. 1046 (Clark) at 2. A person of ordinary skill would have understood that such baits were used to enrich or amplify exome sequences over other sequences in a sample. It was also known that “exome sequencing [described in Clark] can be extended to general enrichment principles and custom enrichment assays. [Clark] demonstrated multiple levels of bait density and genome coverage that can be used as a guide when designing custom enrichment bait sets.” *Id.* at 9; *see also* Section V.C.3.b).

d) “using said probe set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in an additional biological sample from said subject or at least one biological relative of said subject, to generate a sequencing library”

195. Newman discloses using the probe set of step (c) on an additional biological sample of the subject to generate a sequencing library. For example, in Figure 1b discussed previously, Newman discloses that “[l]ater blood draws” can be subjected to the same CAPP-Seq process as the “[f]irst blood draw.” Ex. 1043 at 548.



Newman also discloses in Figure 5d a time course study with multiple serial analyses on cfDNA or tumor DNA for a particular subject, which involve the use of the probe set (*i.e.*, the 157 kb selector in the case of the GBM subject) for an additional biological sample from the subject.



196. A sequencing library is created. Indeed, the “Seq” in CAPP-Seq stands for “deep sequencing” a sequencing library. Ex. 1043 at 1 (Abstract); *see also id.* at “Library preparation and sequencing” in “Online Methods.” For the GBM subject, Newman “applied this selector [*i.e.*, the 157-kb selector for the ~1,500 coding mutations] to perform deep resequencing of the GBM tumor” *Id.* at “ctDNA detection limits for iDES, duplex sequencing, and other methods” in “Online Methods.”

e) “subjecting said sequencing library to sequencing at a coverage of 175 kilobases (kb) or less to identify a presence or absence of at least a subset of said set of genetic variants in said additional biological sample from said subject or said at least one biological relative, wherein a sequencing depth of said sequencing of (e) is greater than a sequencing depth of said sequencing of (a)”

197. The supplementary table shows that the two glioblastoma samples were whole exome sequenced at median depths of 62 and 65 respectively. Ex. 1045 at Table 2, tab entitled “SuppTable2_samples_and_QC”, rows 286-87.

198. Newman discloses a 157 kb selector (capture panel) for the glioblastoma tumor they studied to identify the presence or absence of genetic variants. Ex. 1043 at “ctDNA detection limits for iDES, duplex sequencing, and other methods” in “Online Methods.” Newman “applied this selector to perform deep resequencing of the GBM tumor (1/2 of a MiSeq lane)” *Id.* A person of ordinary skill in the art knew from Illumina literature, among other sources, that a

MiSeq lane produced 4.5-5.1 Gb of sequencing data for 2 x 150 paired end reads. Ex. 1047 at 3 (“MiSeq System Performance Parameters, MiSeq Reagent Kit v2”). Half of a lane would produce 2.25-2.55 Gb of sequencing data. The read depths are therefore calculated to be $(2.25 \times 10^9 \text{ bases} / 157,000 \text{ bases}) = 14,331$ to $(2.55 \times 10^9 \text{ bases} / 157,000 \text{ bases}) = 16,242$.

199. In sum, Newman discloses sequencing depth 62 or 65 for (a), and sequencing depths of 14,331-16,242 for (e). The sequencing depth of (e) is greater than the sequencing depth of (a), as claimed.

6. Claim 2

200. Claim 2 recites “[t]he method of claim 1, further comprising outputting a report that is generated at least based on comparison of results from said sequencing of (a) with results from at least said subjecting said sequencing library to said sequencing of (e).”

201. Newman discloses outputting the report of claim 2. For example, in the example of sequencing a tumor from a GBM patient, Newman performed whole exome sequencing and used the sequencing information to produce a 157-kb personalized selector (capture panel). Ex. 1043 at “ctDNA detection limits for iDES, duplex sequencing, and other methods” in “Online Methods.” Newman “applied this selector to perform deep resequencing of the GBM tumor (1/2 of a MiSeq lane), and confirmed 97% of expected SNVs, 99.7% of which had duplex support (n = 1,502).”

Id. The outputs of 97% of expected SNVs and 99.7% having duplex support constitute reports generated based on comparison of the whole exome sequencing (the sequencing of (a)) to deep resequencing of the tumor (the sequencing of (e)).

7. Claim 3

202. Claim 3 recites “[t]he method of claim 1, wherein in said sequencing of (a), said at least one biological sample is obtained from said subject at a first time point, and wherein said additional biological sample is obtained from said subject or said at least one biological relative at a second time point subsequent to said first time point.”

203. As explained above for claim 1(d), Newman discloses the subject matter of claim 3 for the same reasons.

8. Claim 4

204. Claim 4 recites “[t]he method of claim 1, wherein said sequencing of (a) comprises (i) exome sequencing, (ii) sequencing a panel of genes, (iii) whole genome sequencing, and/or (iv) sequencing a population of complementary deoxyribonucleic acid molecules derived from ribonucleic acid molecules.”

205. As explained above for claim 1(a), Newman discloses the subject matter of claim 4. In particular, among other things, Newman performed whole exome sequencing for the glioblastoma tumor discussed above for claim 1(a).

9. Claim 5

206. Claim 5 recites “[t]he method of claim 1, wherein each of said plurality of nucleic acid probe molecules includes oligonucleotide-directed genomic content comprising (i) at least one variable portion from a result of said sequencing of (a) and (ii) at least one fixed portion independent of said result of said sequencing of (a).”

207. As discussed above for claim 1, Newman discusses a plurality of nucleic acid probe molecules with at least one variable portion from a result of the sequencing of claim 1(a) for analysis of the glioblastoma tumor. The personalized selector used for the glioblastoma analysis contained variable sequences based on the sequencing result of step 1(a).

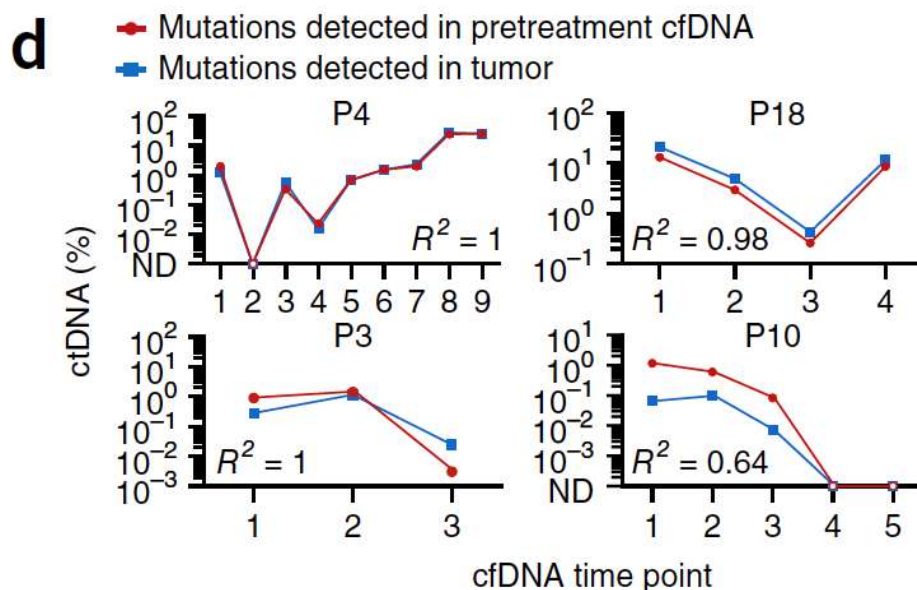
208. In addition, Newman discloses a plurality of nucleic acid probe molecules with at least one fixed portion independent of the result of the step 1(a) sequencing. For example, in its “NSCLC selector design,” Newman taught designing a fixed selector “to target a median of at least eight mutations per NSCLC patient.” *Id.* at “NSCLC selector design” in Online Methods.” Newman also states that “additional clinically relevant regions were included ... to target actionable fusions in *ALK*, *ROS1*, and *RET* using intronic probes, and to ensure coverage of selected regions containing other actionable variants.” *Id.* All of the contents of these

NSCLC selector panels are nucleic acid molecules with fixed sequences that are independent of the sequencing results of step 1(a).

10. Claim 9

209. Claim 9 recites “[t]he method of claim 1, further comprising outputting a report that identifies said presence or absence of said at least said subset of said genetic variants in said subject or said at least one biological relative.”

210. As explained above for claim 1(d), Figure 5d in Newman is a report, in the form of a time course plot, identifying the presence (results above “ND,” or not detected) or absence (“ND”) of genetic variants (mutations) in the subject.



11. Claim 10

211. Claim 10 recites “[t]he method of claim 1, further comprising providing a therapeutic intervention at least based on said presence or absence of said at least said subset of said genetic variants identified in (e).”

212. Newman discloses that “analysis of ctDNA is likely to play a major role in personalized cancer therapy by addressing several unmet clinical needs, including (i) noninvasive genotyping of clinically actionable tumor-derived variants *to guide therapy selection*, (ii) detection of emergent, actionable variants associated with *resistance during targeted therapy*, (iii) early measurement of *therapeutic responses*, and (iv) *detection of minimal residual disease* that is radiographically occult during disease surveillance.” Ex. 1043 at 554. Newman’s analysis of ctDNA therefore has numerous types of therapeutic interventions based on the presence or absence of genetic variants. The ‘611 patent makes clear that “[t]he therapeutic intervention may be a treatment or a report.” Ex. 1001 at 31:20-21.

12. Claim 11

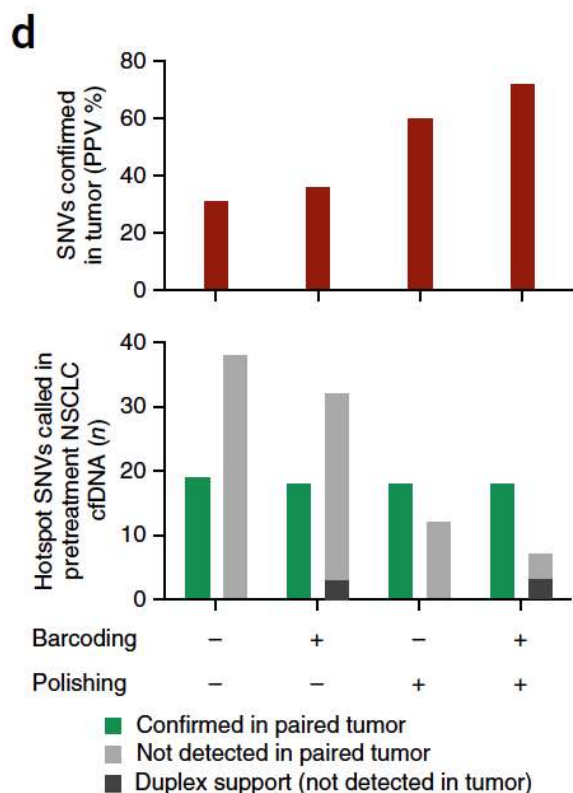
213. Claim 11 recites “[t]he method of claim 1, wherein said at least one biological sample includes a tumor sample and said nucleic acids molecules are from cells in said tumor sample.”

214. Newman discloses the following passage that includes analysis of lung tumor samples:

To characterize the performance of our molecular barcoding strategy, we applied a redesigned CAPP-Seq selector (i.e., capture panel) to maximize the number of mutations per NSCLC patient while minimizing panel size and sequencing cost. The resulting panel was expected to yield at least eight mutations per tumor in 50% of NSCLC patients (Supplementary Table 1). We confirmed the expected performance of this panel by applying it to formalin-fixed, paraffin-embedded (FFPE) tumor specimens and peripheral blood leukocyte samples from NSCLC patients (Supplementary Tables 1–3).

Ex. 1043 at 548.

215. Newman also “assessed iDES-enhanced CAPP-Seq genotyping using 66 serial plasma samples from patients with NSCLC, all of which had tumor biopsies available.” *Id.* at 551. Figure 4d shows the results, including results from the tumor biopsies.



13. Claim 13

216. Claim 13 recites “[t]he method of claim 1, wherein said plurality of nucleic acid sequences having said set of genetic variants comprises genetic variants in a germline sequence of said subject.”

217. In its Supplementary Table 2, Newman reports a “Sample type” of “PBL,” which stands for peripheral blood lymphocyte. Ex. 1045 at Table 2, tab entitled “SuppTable2_samples_and_QC”, rows 287. PBLs contain germline sequence, and Newman’s sequencing of PBLs resulted in a plurality of nucleic acid sequences having said set of genetic variants comprising genetic variants in a germline sequence of said subject.

14. Claim 18

218. Claim 18 recites “[t]he method of claim 1, wherein subjecting said sequencing library to sequencing in (e) comprises generating nucleic acid sequence data from said additional biological sample.”

219. As discussed above for Claim 1(d), Newman discloses using an additional biological sample to generate a sequencing library, which is then used to generate sequence data in step 1(e).

15. Claim 19

220. Claim 19 recites “[t]he method of claim 1, wherein computer processing of said sequence data comprises identifying one or more of (i) genetic variants of said nucleic acid sequence with respect to a germline sequence(s), (ii) alleles which match said reference sequence(s) and are correlated with a type of cancer or other disease, (iii) alleles which determine a human leukocyte antigen (HLA) type, (iv) metrics of gene expression or allele-specific expression, and (v) quantification of non-coding ribonucleic acid (RNA) molecules or micro-RNA molecules which are at least partially tissue-type specific or cancer-type specific.”

221. Newman discloses computer processing of the sequence data by identifying genetic variants of the nucleic acid sequence with respect to a germline sequence. In particular, Newman’s study of a glioblastoma tumor to find ~1500 coding mutations involved comparison of the tumor sequences to germline

sequences in peripheral blood lymphocytes (PBLs), as shown in Supplementary Table 2 for the GBM entries (samples 286-87). Id. at “ctDNA detection limits for iDES, duplex sequencing, and other methods” in “Online Methods”; Ex. 1045 at Sheet 2 (“Sample inventory with QC metrics”), rows 286-87.

16. Claim 28

222. Claim 28 recites “[t]he method of claim 1, further comprising using said presence or absence of said at least said subset of said set of genetic variants identified in (e) to treat said subject or said at least one biological relative for a disorder.”

223. As discussed for claim 10, Newman discloses using identified genetic variants to treat the subject.

17. Claim 29

224. Claim 29 recites “[t]he method of claim 1, wherein said reference comprises a human reference sequence.”

225. As discussed above for claim 19, Newman discloses the use of reference germline sequences from PBLs, which is a human reference sequence.

18. Claim 30

226. Claim 30 recites “[t]he method of claim 1, wherein said reference is a human reference sequence.”

227. Newman discloses claim 30 for the same reasons as claim 29 above.

C. **Ground 4: Newman Renders Obvious Claims 1-5, 9-11, 13, 18, 19, 28-30**

228. Claims 1-5, 9-11, 13, 18, 19, and 28-30 would have been obvious over Newman (Ex. 1043), as discussed for Ground 3 and further discussed below for Ground 4.

229. Claim 1 recites:

A method for personalized genetic testing, comprising:

(a) sequencing nucleic acid molecules from at least one biological sample of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more;

(b) computer processing said sequencing data to identify a plurality of nucleic acid sequences having a set of genetic variants, wherein said set of genetic variants are identified with respect to a reference;

(c) obtaining a probe set comprising a plurality of nucleic acid probe molecules having said plurality of nucleic acid sequences or complements thereof, wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in said at least one biological sample;

(d) using said probe set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in an additional biological sample from said subject or at least one biological relative of said subject, to generate a sequencing library; and

(e) subjecting said sequencing library to sequencing at a coverage of 175 kilobases (kb) or less to identify a presence or absence of at least a subset of said set of genetic variants in said additional biological sample from said subject or said at least one biological relative, wherein a sequencing depth of said sequencing of (e) is greater than a sequencing depth of said sequencing of (a).

1. Claims 1-5, 9-11, 13, 18, 19, 28-30

230. To the extent claims 1-5, 9-11, 13, 18, 19, or 28-30 are not anticipated as explained in Ground 3, they would have been obvious in view of Newman.

231. To the extent Newman's disclosure does not anticipate the limitation in step 1(a) that the sequencing comprises 12 gigabases or more, then a person of ordinary skill would have found it obvious to sequence 12 gigabases or more based on Newman's disclosure.

232. As previously discussed, Newman discloses that "[s]equencing was performed using 2 x 100 or 2 x 150 paired-end reads" Ex. 1043 at "Library preparation and sequencing" in "Online Methods." In its supplementary tables, published online together with and at the same time as the primary Newman article, Newman discloses data from the sequencing of two samples from a glioblastoma patient. Ex. 1045 at Table 2, tab entitled "SuppTable2_samples_and_QC", rows 286-87. In particular Newman reports that these two sequencing runs resulted in totals of 82,247,848 and 86,781,192 reads respectively. *Id.*

Sample number	Cohort	Case	Collection time ^a	Sample type	Input mass (ng)	Selector ^b	Sequencing adapter ^c	De-duplication ^d	Total reads
275	Healthy/Te	GBM-CTR25	GBM-CTR25-	cfDNA	32.0	GBM	Tandem	Barcoding	37,647,656
276	Healthy/Te	GBM-CTR26	GBM-CTR26-	cfDNA	32.0	GBM	Tandem	Barcoding	38,321,932
277	Healthy/Te	GBM-CTR27	GBM-CTR27-	cfDNA	32.0	GBM	Tandem	Barcoding	33,065,346
278	Healthy/Te	GBM-CTR28	GBM-CTR28-	cfDNA	32.0	GBM	Tandem	Barcoding	35,976,502
279	Technical	GBM-Spike-0.	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	103,288,518
280	Technical	GBM-Spike-0.	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	105,786,270
281	Technical	GBM-Spike-0.	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	100,222,542
282	Technical	GBM-Spike-0.	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	66,777,970
283	Technical	GBM-Spike-0.	GBM-Spike-C	cfDNA	32.0	GBM	Tandem	Barcoding	56,060,804
284	Technical	HD500-0.5%-i	HD500-0.5%-	Cell line bl	72.0	NSCLC tumor ge	Tandem	Barcoding	270,879,108
285	Healthy/Te	CTR31	CTR31-c1	cfDNA	32.0	NSCLC tumor ge	Tandem	Barcoding	259,725,374
286	GBM	GBM Relapse	na	Tumor	100.00	SeqCap EZ Hum.	Tandem	Barcoding	82,247,848
287	GBM	GBM PBL	na	PBL	100.00	SeqCap EZ Hum.	Tandem	Barcoding	86,781,192

Using the 100 bp read length, the sequencing comprises 8.22 GB and 8.68 GB respectively. Also, the claim language instructs that one may combine the two samples (“sequencing nucleic acid molecules from *at least one biological sample* of a subject to generate sequencing data, wherein said sequencing comprises sequencing 12 gigabases (GB) or more”), which would result in sequencing comprising 8.22 GB + 8.68 GB = 16.9 GB.

233. A POSA would have been motivated to modify Newman so that, even with 100 bp read length and no combination of values (*i.e.*, the 8.22 GB and 8.68 GB values above, but not 16.9 GB), the sequencing would be done with total reads that result in sequencing of 12 GB or more. As previously discussed, the ’611 patent does not ascribe any inventiveness to sequencing 12 gigabases or more. The ’611 patent discusses 12 gigabases only once in an example—Example 2—by stating that “a germline DNA sample can be exome sequenced (e.g., 12 billion bases).” Ex. 1001 at 62:33-35. The use of “e.g.” would have been understood by a person of ordinary

skill to mean that 12 billion bases (gigabases) was simply an estimate of how much sequencing would probably be required for the '611 patent method, and a person of ordinary skill would have also understood and been motivated by the fact that more sequencing data would provide greater accuracy of the results.

234. A POSA would have reasonably expected to succeed in modifying Newman as described above. There is nothing in Newman suggesting that sequencing of 12 GB, for example, would not succeed. As discussed previously in the background discussion of sequencing depth and read counts (Section V.C.2), Sims disclosed prior art reports of whole exome sequencing in a broad range was about 0.2 GB to 20 GB. 12 GB of sequencing obviously lies in the heart of the exome sequencing range in Sims. Therefore, a POSA would have reasonably expected to achieve success by sequencing 12 GB or more for the glioblastoma samples rather than 8 GB.

2. Claim 5—Further Information Regarding Obviousness

235. Claim 5 recites “[t]he method of claim 1, wherein each of said plurality of nucleic acid probe molecules includes oligonucleotide-directed genomic content comprising (i) at least one variable portion from a result of said sequencing of (a) and (ii) at least one fixed portion independent of said result of said sequencing of (a).”

236. Newman discloses a plurality of nucleic acid probe molecules with at least one fixed portion independent of the result of the step 1(a) sequencing. For example, in its “NSCLC selector design,” Newman taught designing a fixed selector “to target a median of at least eight mutations per NSCLC patient.” Ex. 1043 at “NSCLC selector design” in Online Methods.” Newman also states that “additional clinically relevant regions were included ... to target actionable fusions in *ALK*, *ROS1*, and *RET* using intronic probes, and to ensure coverage of selected regions containing other actionable variants.” *Id.* All of the contents of these NSCLC selector panels are nucleic acid molecules with fixed sequences that are independent of the sequencing results of step 1(a).

237. There is no reason given in the ’611 patent why fixed sequences such as these could not be combined with variable sequences such as the ones used for the GBM patient’s tumor, and a person of ordinary skill in the art knew that a panel of 175 kb or less could be designed and produced with fixed and variable sequences.

238. In addition to the PCR probes comprising the fixed portion already included in Newman’s NSCLC selector, a POSA reading Newman in view of the knowledge in the field would have also found it obvious to include probes targeted to genetic variants involved in the pharmacogenomics of cancer drugs, *i.e.*, genetic variants that were known to have companion diagnostics, as well as to tumor-driver mutations that were known in the art.

239. For example, as explained in Section V.F (“Providing Actionable Information to Healthcare Providers”) above, many effective cancer therapies are targeted to particular genetic mutations that present in certain forms of cancer. Examples include the blockbuster drug Herceptin targeting HER2 positive breast cancer (*i.e.*, forms of breast cancer carrying mutations in the HER2 gene), which was first introduced in 1998. *See* Ex. 1014 (Ross 2011) at 527 (“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.”).

240. Relatedly, the identity of therapy/cancer pairs would have been readily ascertainable by one skilled in the art by accessing the treatment guidelines for a particular cancer promulgated by the National Comprehensive Cancer Network (called the NCCN Guidelines) which are widely used by oncologists to reference and define standard of care treatments. Similar information would have also been available from the FDA-approved prescribing information available for relevant

drugs. To design a probe for relevant mutations identified in this manner, a POSA would have been able to determine the relevant sequence information from any one of the publicly available cancer databases, such as COSMIC, *see* Section V.B (“Genetic Mutations”), and then designed nucleic acid probes based on the techniques already known in the art and applied in the methods expressly disclosed by Ley. Accordingly, a POSA would have reasonably expected success in the predictable use of known information (regarding the sequences) and known techniques (the creation of nucleic acid probes to target such sequences) to add “fixed” probes to the types of assays described in Ley in order to provide information relevant to the selection and response to mutation-specific therapy pairings.

241. Additionally, as discussed above, a POSA would have been familiar with known mutations common in certain forms of cancer, which are recorded in databases based on population-level studies. A POSA would have been motivated to combine candidate gene-based probes (fixed portion) in a panel with patient specific mutations discussed above (variable portion) in order to assess the cancer for known mutations in well-characterized and studied candidate genes in addition to patient-specific mutations. For example, as noted above, in many instances such genes may drive the course of the cancer and may be key targets for therapy. Further, such a combination of fixed and variable probes would overcome the limitations of a purely

hypothesis-driven (fixed probe-only) approach, namely missing mutations in a particular patient sample.

242. A POSA would have known that one reason a targeted approach may miss such mutations is because of tumor heterogeneity. Tumor heterogeneity refers to the fact that genetic variants present in a cancer are not always uniformly dispersed across a tumor spatially, and that certain key variants may not be present in the tumor at all times during the cancer's progression. *See, e.g.*, Ex. 1071 (Burrell) at 338 ("Extensive genetic and phenotypic variations exist not only between tumours (intertumour heterogeneity) but also within individual tumours (intratumour heterogeneity; Fig. 1). In tumours, populations of subclones can intermingle or be spatially separated, and this subclonal architecture varies dynamically throughout the disease course."). This explains why sequencing a tumor sample may miss some of the patient's tumor-associated mutations at any given point in time along the course of the disease. As a result, it would have been obvious to include probes for selectively enriching such hotspot mutations (*i.e.*, the claimed fixed probes) in the panel in order to detect in subsequent samples the mutations previously identified via the WGS assay, but also mutations that were known to be common in the patient's type of cancer. A POSA would have been motivated to monitor hotspot mutations for a particular cancer (*i.e.*, the fixed probes) even in cases where they may not have been identified at the time of initial presentation or at the time of an

initial whole-genome sequencing assay, as such mutations may present later in the course of the patient's cancer.

243. A POSA would have been able to readily create probes for such mutations using already known techniques. Indeed, Newman already recognized that such probes were well-known in the art, and thus would have reasonably expected success in including such "fixed" probes in a cancer-monitoring panel. Further, such fixed probes would selectively enrich sequences comprising genetic variants in exactly the same way as variable probes selectively enrich sequences comprising patient-specific genetic variants.

IPR2023-00546
U.S. Patent No. 10,450,611

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: February 10, 2023

By: _____

 Digitally signed by
John Quackenbush
Date: 2023.02.10
13:09:14 -05'00'

John Quackenbush, Ph.D.