UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Foresight Diagnostics Inc.,
Petitioner,

v.

Personalis, Inc.,
Patent Owner.

IPR2023-00545
U.S. Patent No. 11,299,783

DECLARATION OF JOHN QUACKENBUSH, PH.D. IN SUPPORT OF PETITION 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

Page

I. Introduction 1

II. Legal Standards 7

A. Claim Construction 7

B. Anticipation 8

C. Obviousness 8

III. The ’783 Patent 12

IV. Person of Ordinary Skill in the Art 12

V. Overview of a Person of Ordinary Skill in the Art’s Knowledge 14

A. Overview of Molecular Biology and Genetics 14

1. The Central Dogma of Molecular Biology 14

2. Genetics 17

B. Genetic Mutations 19

C. Nucleic Acid Sequencing Technologies 22

1. Sequencing Techniques 22

2. Sequencing Depth and Read Counts 27

3. Amplification and Enrichment 31

D. Personalized Cancer Testing and Monitoring 40

E. Related Applications of Next Generation Sequencing 45

F. Providing Actionable Information to Healthcare Providers 47

TABLE OF CONTENTS
Cont'd

Page

VI. Claim Construction 50

A. "probe" 51

B. "report" 52

VII. Prior Art References of the Grounds 53

A. Ley (Ex. 1012) 53

B. Forshew (Ex. 1030) 54

C. Dawson (Ex. 1070) 54

VIII. The Challenged Claims Are Unpatentable 54

A. Ground 1: Ley Anticipates Claims 1-2, 5-7, 9, 14-15, 17-19, 21, 23-25, and 27 54

1. Claim 1 54

2. Claim 2 73

3. Claim 5 74

4. Claim 6 74

5. Claim 7 75

6. Claim 9 75

7. Claim 14 76

8. Claim 15 76

9. Claim 17 76

10. Claim 18 77

TABLE OF CONTENTS
Cont'd

Page

11. Claim 19 78

12. Claim 21 78

13. Claim 23 79

14. Claim 24 79

15. Claim 25 80

16. Claim 27 81

B. Ground 2: Obviousness Over Ley in view of a POSA's Knowledge 82

1. Claim 1 82

2. Claim 2 90

3. Claim 3 91

4. Claim 4 92

5. Claim 8 93

6. Claim 13 94

7. Claim 14 95

8. Claim 15 95

9. Claim 16 96

10. Claim 18 97

11. Claim 22 98

12. Claim 26 100

TABLE OF CONTENTS
Cont'd

Page

13. Claim 28 101

14. Claim 29 102

C. Ground 3: Obviousness over Forshew in view of Dawson 104

1. Claim 1 104

2. Claim 2 120

3. Claim 3 121

4. Claim 4 122

5. Claim 5 122

6. Claim 6 124

7. Claim 7 124

8. Claim 8 124

9. Claim 9 125

10. Claim 13 125

11. Claim 14 126

12. Claim 15 126

13. Claim 16 127

14. Claim 17 127

15. Claim 18 127

16. Claim 19 128

17. Claim 21 129

TABLE OF CONTENTS
Cont'd

Page

18. Claim 22 129

19. Claim 23 131

20. Claim 24 133

21. Claim 25 133

22. Claim 26 133

23. Claim 27 135

24. Claim 28 135

25. Claim 29 136

IX. Conclusions 137

Exhibits Cited

Ex. 1001	U.S. Patent No. 11,299,783 to West *et al.*
Ex. 1002	Rebecca J. Leary *et al.*, "Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing," *Science Translational Medicine* (2010) (***Leary***)
Ex. 1003	Quake *et al.*, WO 2011/057061 (***Quake***)
Ex. 1004	Matthew W. Anderson & Iris Schrijver, "Next Generation DNA Sequencing and the Future of Genomic Medicine," *Genes* (2010) (***Anderson***)
Ex. 1005	Andreas Gnirke *et al.*, "Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing," *Nature Biotechnology* (2009) (***Gnirke***)
Ex. 1006	Florian Mertes *et al.*, "Targeted enrichment of genomic DNA regions for next-generation sequencing, Briefings in Functional Genomics," *Briefings in Functional Genomics* (2011) (***Mertes***)
Ex. 1007	Lira Mamanova *et al.*, "Target-enrichment strategies for next-generation sequencing," *Nature* (2010) (***Mamanova***)
Ex. 1008	K.C. Allen Chan *et al.*, "Cancer Genome Scanning in Plasma: Detection of Tumor-Associated Copy Number Aberrations, Single-Nucleotide Variants, and Tumoral Heterogeneity by Massively Parallel Sequencing," *Clinical Chemistry* (2013) (***Chan***)
Ex. 1009	Gary J.W. Liao *et al.*, "Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles," *Clinical Chemistry* (2011) (***Liao***)
Ex. 1010	*Personalis v. Foresight*, First Amended Complaint, D. Colo., Case No. 1:22-v-01913, Dkt. 9, filed August, 17, 2022 ("FAC")
Ex. 1011	Reserved

Ex. 1012	Timothy J. Ley *et al.*, "DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome," *Nature* (2008) (***Ley***)
Ex. 1013	Jay Shendure & Hanlee Ji, "Next-generation DNA sequencing," *Nature Biotechnology* (2008) (***Shendure***)
Ex. 1014	Jeffrey S. Ross and Maureen Cronin, "Whole Cancer Genome Sequencing by Next-Generation Methods," *American Journal of Clinical Pathology* (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," *Genes, Chromosomes & Cancer* (2010) (***McBride***)
Ex. 1016	David M. Kurtz, et al., "Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA," *Nature Biotechnology* (2021) (***Kurtz***) (Ex. J to FAC)
Ex. 1017	https://www.aacc.org/science-and-research/clinical-chemistry/clinical-chemistry-podcasts/2013/cancer-genome-scanning-in-plasma-detection-of-tumor-associated-copy-number-aberrations
Ex. 1018	NIH Spreadsheet Regarding DNA Sequencing Costs
Ex. 1019	https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
Ex. 1023	*Curriculum Vitae* 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	Seo Young Oh, et al., "Improved DNA Extraction Method for Molecular Diagnosis from Smaller numbers of Cells," *Korean Jornal of Clinical Laboratory Science* (2014) (***Oh***)
Ex. 1026	Leary Supplemental Information

Ex. 1027	Illumina, Technical Note: Sequencing
Ex. 1028	Illumina, Specification Sheet: Illumina Sequencing
Ex. 1029	Hugo Lam, et al., "Performance comparison of whole-genome sequencing platforms," *Nature Biotechnology* (2012) (***Lam***)
Ex. 1030	Tim Forshew, et al., "Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA," *Science Translational Medicine* (2012) (***Forshew*)**
Ex. 1031	Matthew Meyerson, et al., "Advances in understanding cancer genomes through second-generation sequencing," *Nature Reviews* (2010) (***Meyerson*)**
Ex. 1032	Reserved
Ex. 1033	Reserved
Ex. 1034	Reserved
Ex. 1035	Reserved
Ex. 1036	Reserved
Ex. 1037	Reserved
Ex. 1038	Reserved
Ex. 1039	Reserved
Ex. 1040	D. Sims *et al.*, "Sequencing depth and coverage: key considerations in genomic analyses," Nature Reviews: Genetics (2014) (***Sims***)
Ex. 1041	H. Li, "Mapping short DNA sequencing reads and calling variants using mapping quality scores," Genome Res. (2008) (***Li***)
Ex. 1042	Ley Supplemental Information

Ex. 1043	A. Newman *et al.*, "Integrated digital error suppression for improved detection of circulating tumor DNA," Nature Biotechnology (2016) (***Newman***)
Ex. 1044	A. Newman *et al.*, "An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage," Nature Medicine (2014) (***Newman 2014***)
Ex. 1045	Newman Supplementary Excel Spreadsheet
Ex. 1046	M. Clark *et al.*, "Performance comparison of exome DNA sequencing technologies," Nature Biotechnology (2011) (***Clark***)
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. 1070	S. Dawson, *et al.*, "Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer," New England Journal of Medicine (2013) (***Dawson***)

Ex. 1071	R. Burrell, *et al*., "The causes and consequences of genetic heterogeneity in cancer evolution," Nature (2013) (***Burrell***)
Ex. 1072	Plaintiff Personalis' Infringement Contentions for United States Patent 11,299,783
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) (***Misawa***)
Ex. 1075	S. Davies, *et al.*, "Indications for Hematopoietic Cell Transplantation in Acute Leukemia," Biology of Blood and Marrow Transplantation (2008) (***Davies***)
Ex. 1076	B. Alter, *et al.*, "Clinical and molecular features associated with biallelic mutations in *FANCD1/BRCA2*," Journal of Medical Genetics (2007) (***Alter***)
Ex. 1077	D. Bentley, *et al.*, "Accurate whole human genome sequencing using reversible terminator chemistry," Nature (2008) (***Bentley***)

I. Introduction

1. My name is John Quackenbush. I have been retained by Foresight Diagnostics Inc. ("Petitioner") to provide my independent evaluation and opinions regarding U.S Patent No. 11,299,783 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 '783 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. I have published more than 300 peer reviewed scientific papers that have been cited more than 79,000 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.

16. 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

17. 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

18. 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.

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

20. 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.

21. As discussed below in Section VI ("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, 1011.

B. Anticipation

22. 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

23. First, 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 and not obvious in light of what came before it. That which came before is generally referred to as "prior art."

24. 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.

25. 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 negatived by the manner in which the invention was made.

35 U.S.C. § 103(a).

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. 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 '783 Patent

32. The '783 patent (Ex. 1001) is titled "Methods and Systems for Genetic Analysis." 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 provision application in fact provides sufficient support for the issued claims of the '783 patent.[1]

IV. Person of Ordinary Skill in the Art

33. 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.

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

34. The relevant technology field for the ’783 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 ’783 patent before May 27, 2016.

35. 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.

36. 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

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

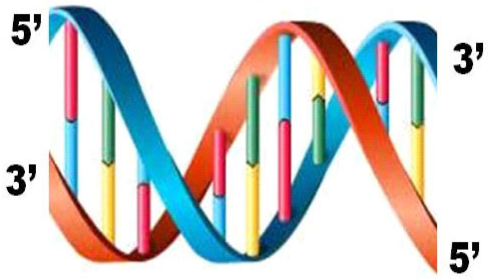
A. Overview of Molecular Biology and Genetics

38. 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

39. 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.

40. 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.

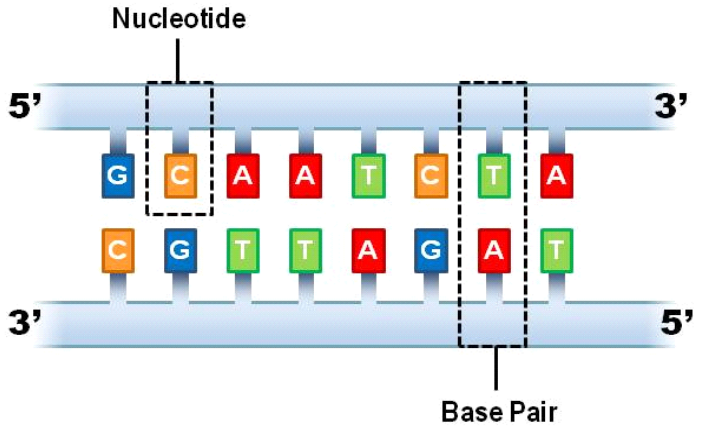


41. The smallest functional unit of DNA is the gene, which is a sequence of nucleotides that encode for ribonucleic acid ("RNA"), which in turn codes for a protein. Proteins are incredibly versatile large molecules that serve crucial functions in essentially all biological processes. As examples: some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

42. 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."

43. 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.

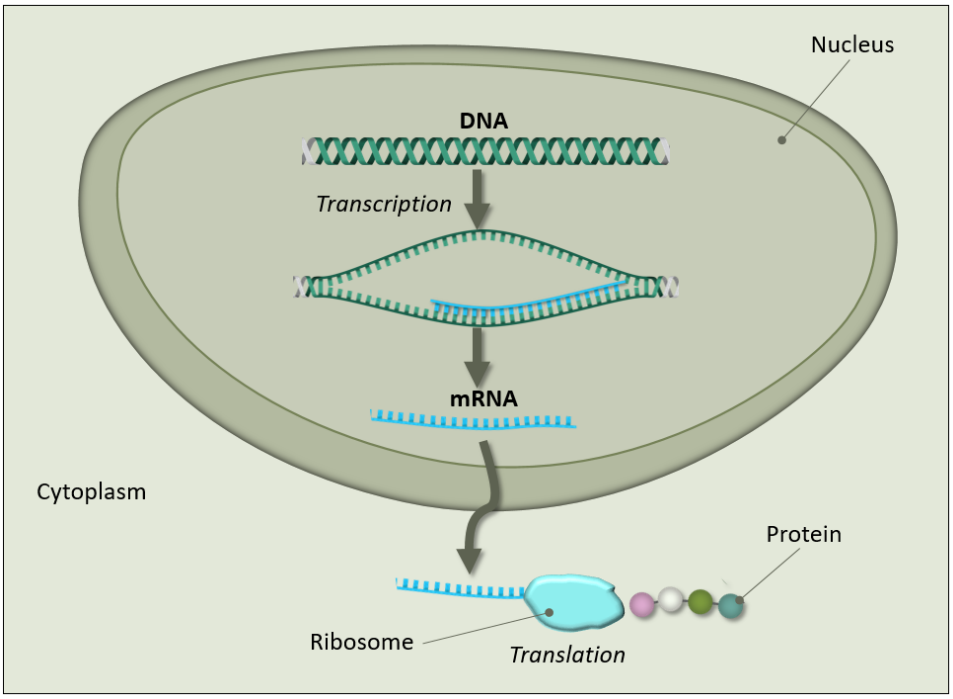


44. The A-T and G-C interactions between strands of oligonucleotides are themselves specific. However, it is possible for two strands to bind together even if not all of the bases in them are perfectly complementary. That is, as long as a sufficient number of A-T and G-C pairs can form between two oligonucleotides,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

45. 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.

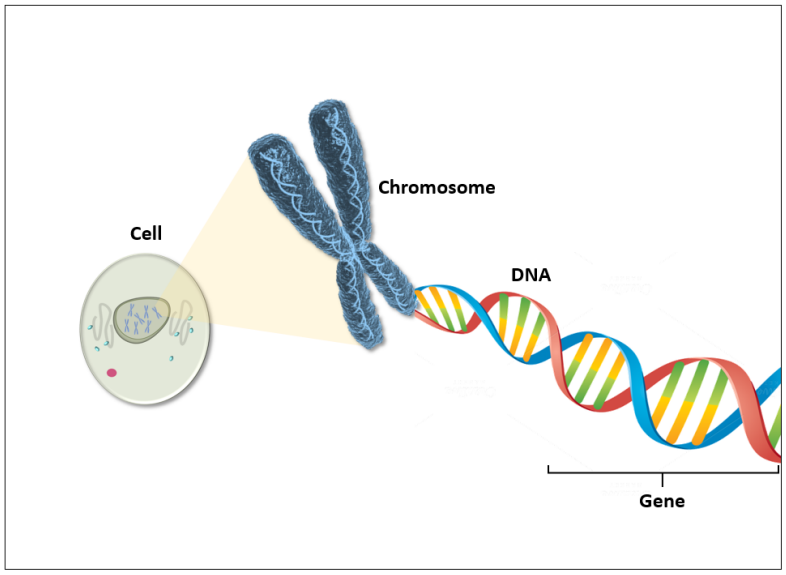
46. The flow of information from DNA to RNA to protein is referred to as "the central dogma of molecular biology."



2. Genetics

47. 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.



48. 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."

49. 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.

50. 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.

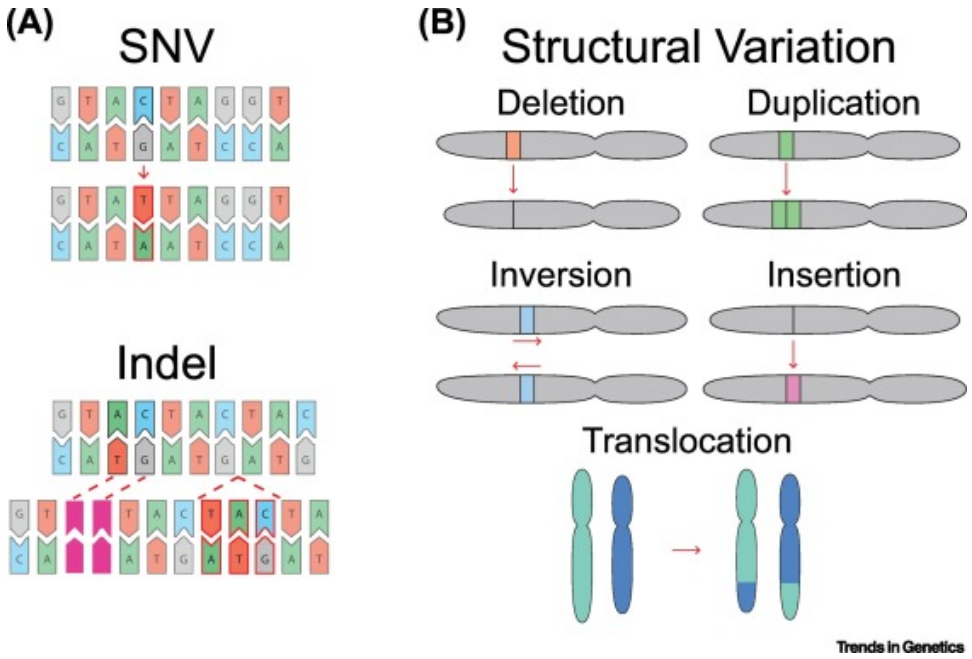
B. Genetic Mutations

51. 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.

52. Different types of somatic mutations were known in the prior art well before January 2013. *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.



53. 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.

54. By well before 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 Cancer Genome Atlas ("TCGA"), International Agency for Research on Cancer ("IARC"), and the Catalogue of Somatic Mutations in Cancer ("COSMIC").

C. Nucleic Acid Sequencing Technologies

1. Sequencing Techniques

55. 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.

56. 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.

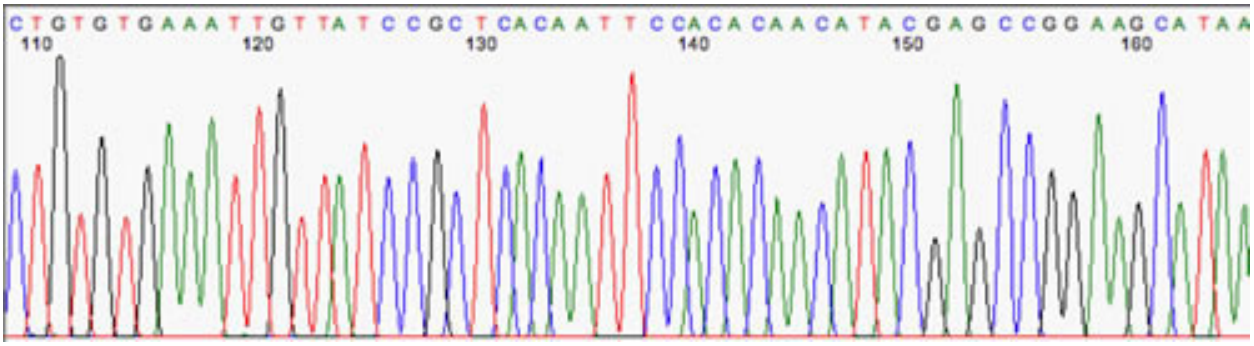
57. In Sanger sequencing, single-stranded DNA "templates" (*i.e.*, the DNA to be sequenced) are placed in solution containing a polymerase, oligonucleotide

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.

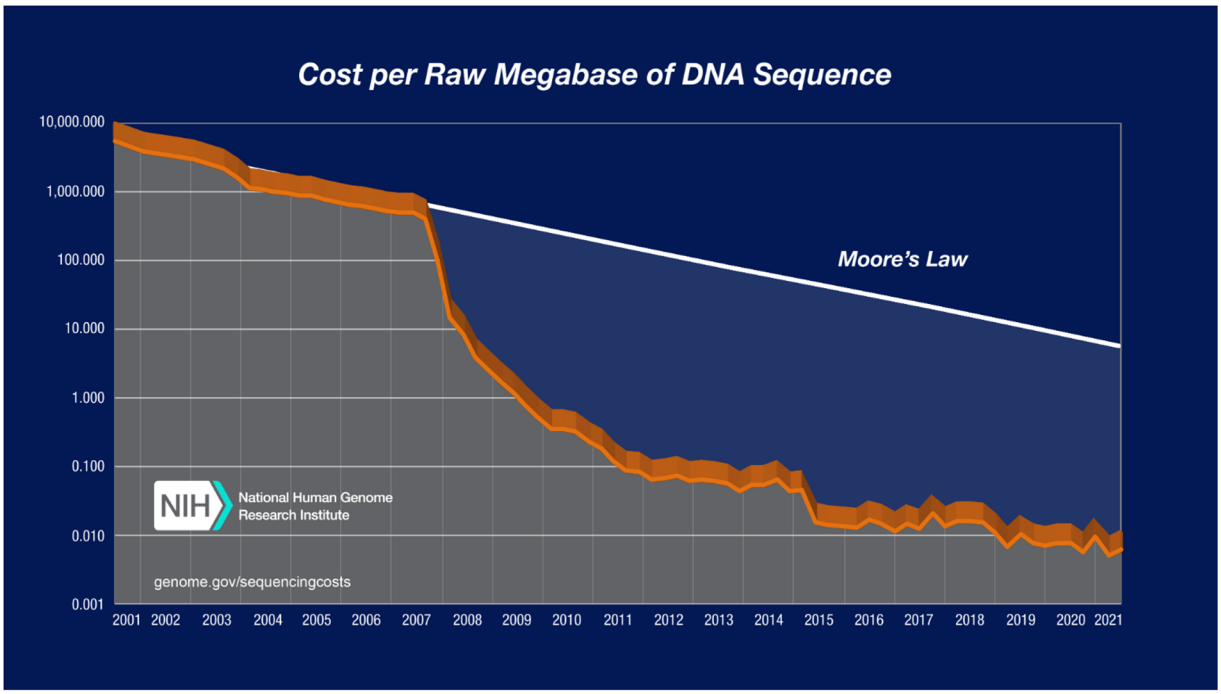
58. During strand extension in Sanger sequencing, when a normal deoxynucleotide is added to the growing strand, the next position remains "open" for another deoxynucleotide or di-deoxynucleotide to pair up to the template and elongate the copy strand. Once a di-deoxynucleotide is incorporated into the 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.

59. 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.



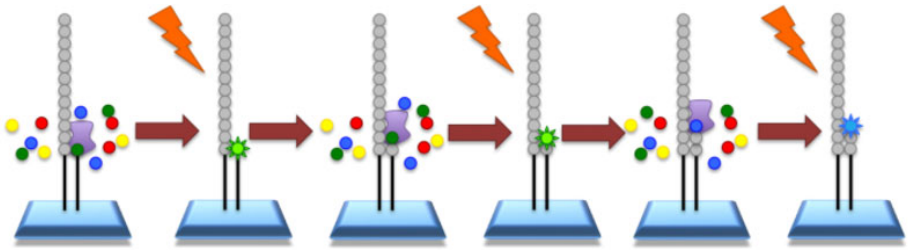
60. By 2005, a new approach to sequencing often referred to as "next generation sequencing" ("NGS"), sometimes also referred to as "massively parallel sequencing," or "sequencing by synthesis" had been developed to allow "multiplex" sequencing of hundreds of thousands or millions of different DNA molecules 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).

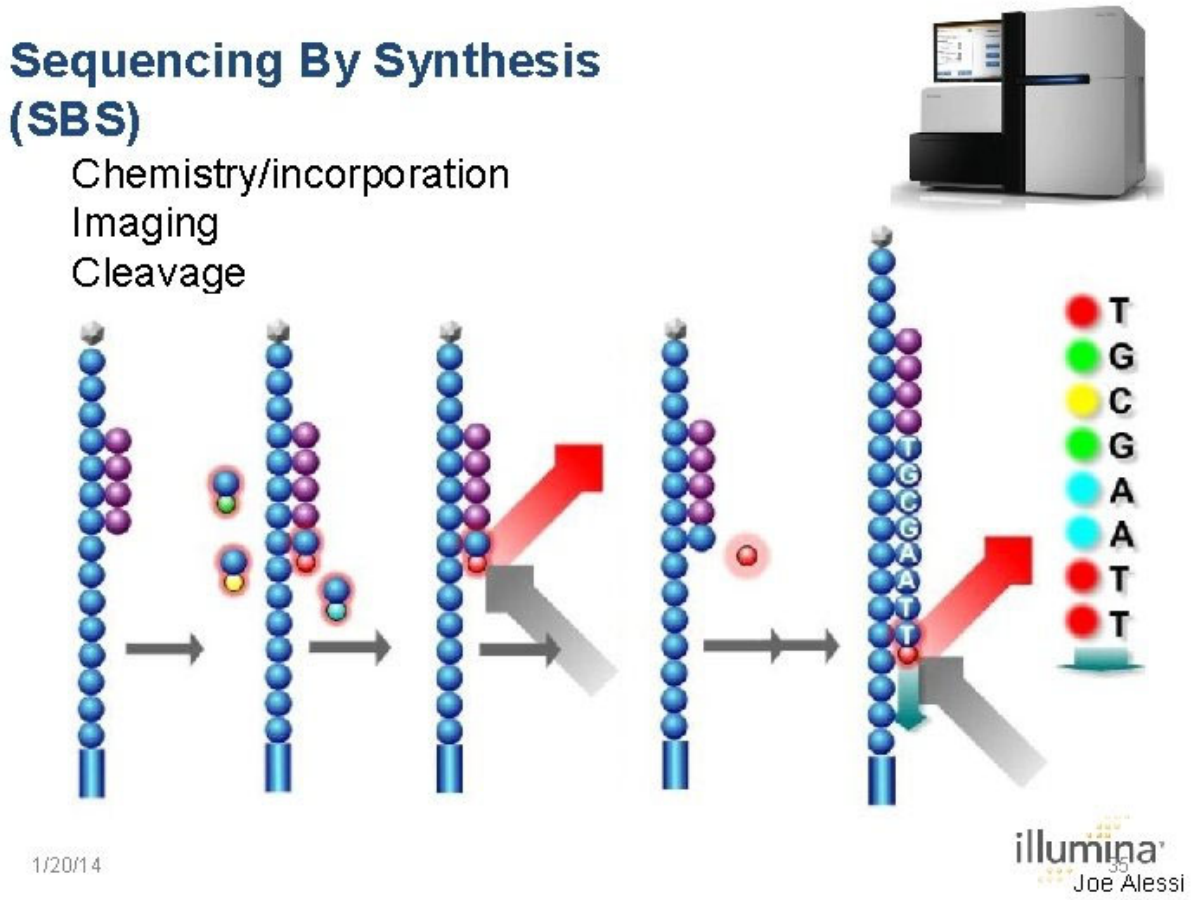


Sequencing cost per megabase - 2021

61. 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.





62. 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

63. 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").

64. "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 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 (Bentley). 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. 1028 (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.

65. 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 HiSeq 2000 and a different sequencing platform from Complete Genomics. Ex. 1029 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. 1027 at 2.

66. 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.

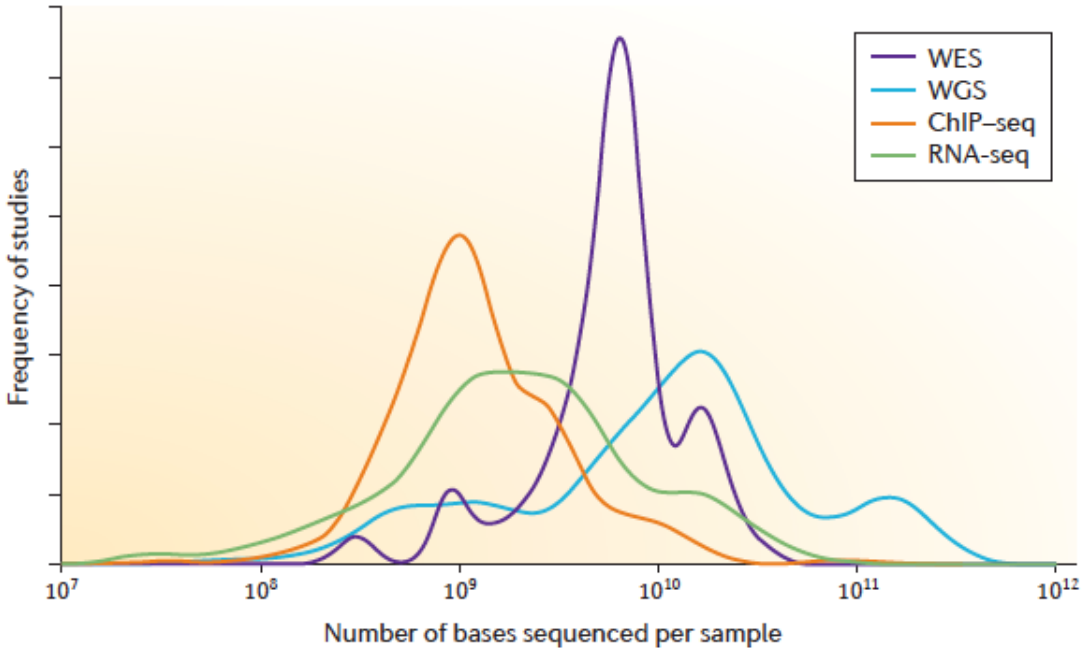


Figure 1 | **Sequencing depths for different applications.** The frequency of studies that use read counts of all runs (which are typically flow-cell lanes) and that were deposited from 2012 to June 2013 for the Illumina platform in the European Nucleotide Archive (ENA) is shown. The plot provides an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

67. 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.

68. 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.

69. 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).

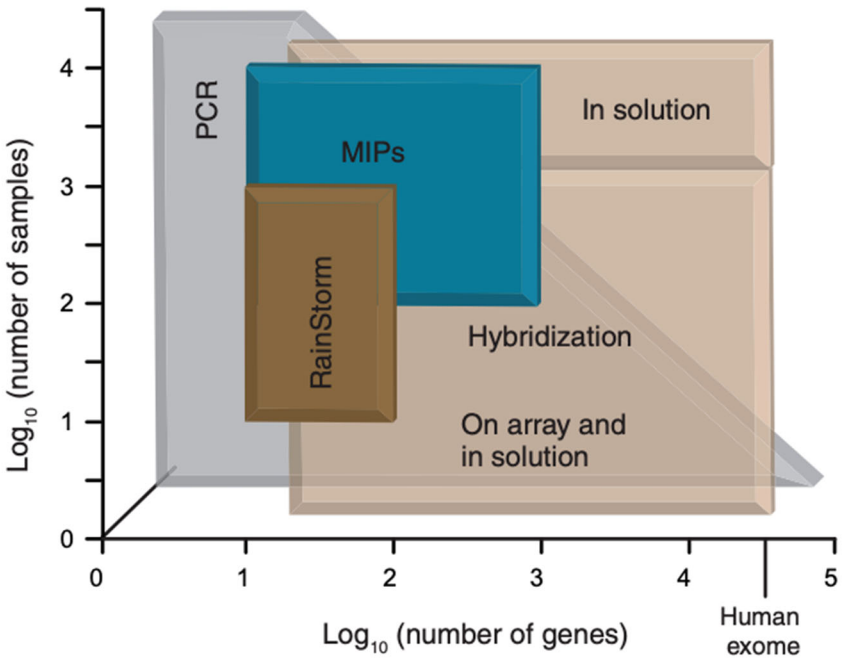


Figure 2 | Suitability of different target-enrichment strategies to different combinations of target size and sample number. Suitability was estimated from the perspective of the feasibility with which each method could be applied to the various combinations of target size and sample number, rather than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

70. While these strategies have considerable overlap in their possible applications, a fair summary of the advantages of each technique is set forth in Ex. 1006 (Mertes 2011) :

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

Ex. 1006 (Mertes 2011) at 380. Thus, in assays where a POSA sought to analyze many sequences of interest, solution-based hybrid capture would have been an ideal enrichment strategy. On the other hand, in assays where a POSA intended to analyze relatively few sequences of interest, then a PCR approach would have been particularly well-suited. For intermediate numbers of potential targets, either technique might serve equally well. However, these parameters would not have been considered as strict limitations, and the techniques can and were used on a relatively interchangeable basis in many experimental protocols, and could also be used together as complements to one another. *See* Ex. 1006 (Mertes 2011) (Table 1)

(describing pros, cons, and overlapping applications of "[c]urrently employed targeted enrichment techniques").

Table 1: Currently employed targeted enrichment techniques

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

All major targeted enrichment techniques show relative pros and cons.

a) PCR

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

72. 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.

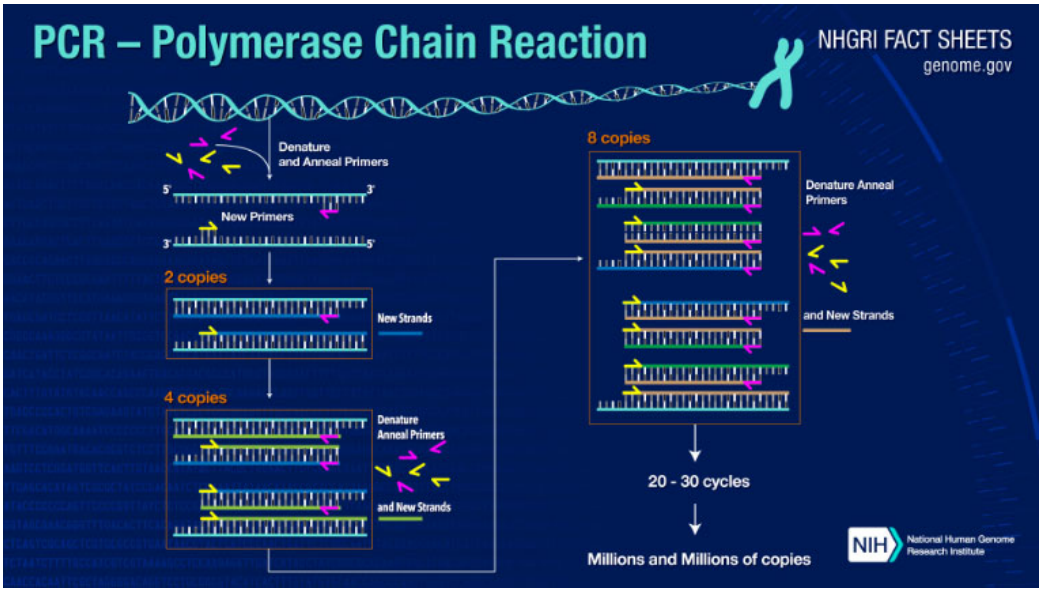
73. PCR "primers" or "probes," which are short oligonucleotides that are complementary to a small portion of the DNA templates, can then be added to the denatured DNA at high concentrations. These primers/probes then hybridize (*i.e.*, bind through base pairing interactions) to the DNA template strand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

74. 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.

75. 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."

76. 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

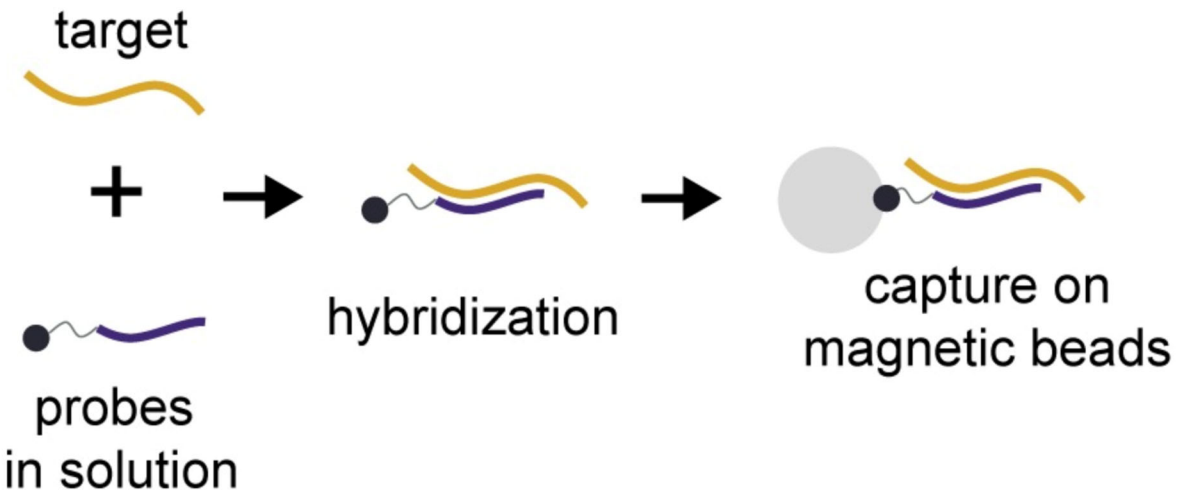
77. A POSA would have also been familiar with the use of solution-based hybrid capture techniques for target enrichment. In short, this technique works by mixing in solution the target DNA with an abundance of short nucleic acid "baits," which hybridize to the target DNA. The baits also include an additional 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).

78. 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."); 374-375 ("Targeted enrichment can be useful in a number of situations where particular portions of a whole genome need to be analyzed.").

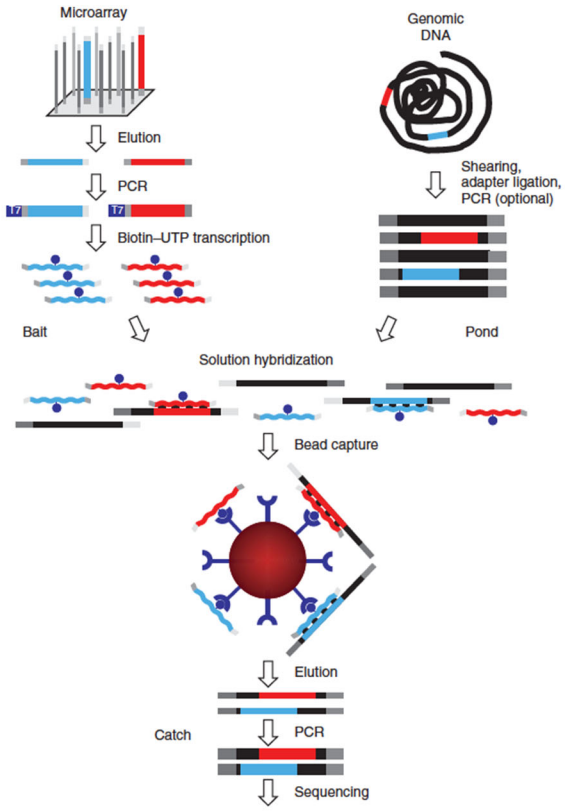


Figure 1 Overview of hybrid selection method. Illustrated are steps involved in the preparation of a complex pool of biotinylated RNA capture probes (bait; top left), whole-genome fragment input library (pond; top right) and hybrid-selected enriched output library (catch; bottom). Two sequencing targets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

79. 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 2011) at 379.

D. Personalized Cancer Testing and Monitoring

80. 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.*

81. 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 2008) 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.*

82. 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.

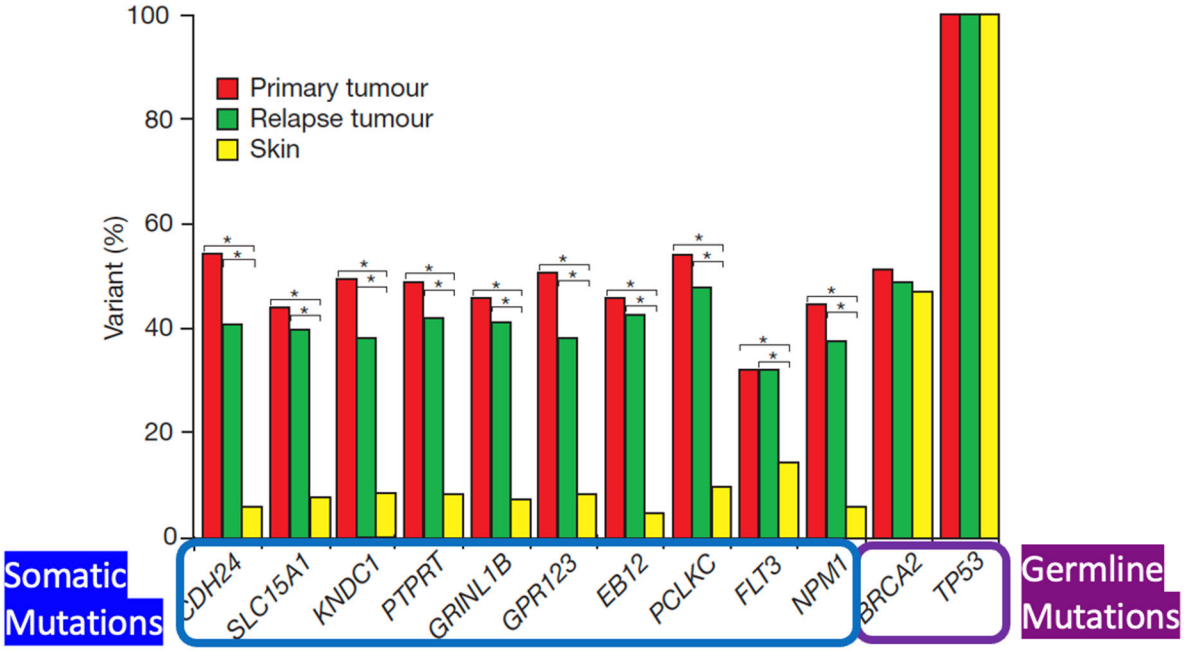


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that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

83. 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).

84. 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

85. 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.

86. 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.

87. 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.

88. 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

89. 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.

90. 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").

91. "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.

92. 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. Claim Construction

93. have generally applied the plain and ordinary meaning that the claims of the ’783 patent would have had to a POSA in view of the ’783 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 of the ’783 patent:

94. **Claim 1**: A method for personalized genetic testing, comprising:

> (a) obtaining a personalized probe set for a subject, which personalized probe set comprises a plurality ***of nucleic acid probe molecules*** having a 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 a set of one or more genetic variants over other sequences in at least one biological sample of said subject, wherein said set of genetic variants are identified in sequence data for said at least one biological sample, further wherein said personalized probe set comprises nucleic acid probe molecules having oligonucleotide-directed genomic content comprising: (i) at least one variable portion from a result of said sequence data and (ii) at

least one fixed portion independent of said result of said sequencing data

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

(c) subjecting said sequencing library to sequencing to identify a presence or absence of at least a subset of said set of genetic variants in at least said additional biological sample from said subject, wherein a sequencing footprint of said sequencing of (c) is less than a sequencing footprint of said sequence data of (a); and

(d) generating a ***report*** identifying (i) a presence or absence of a health condition or disease of the subject in said additional biological sample based on said identified at least a subset of (c) and (ii) genetic variants which inform a therapy choice or a change in therapy for the subject.

A. "probe"

95. I have applied the meaning the term "probe" would have had to a POSA in view of the '783 patent's specification. The '783 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 4:60-61; s*ee also id*. at 7:25-27 (probes include "synthesizing the plurality of nucleic acid probe molecules on a single solid substrate."). The '783 patent also explains that "nucleic

acid probe molecules comprise primers for amplifying the nucleic acid sequences." *Id*. at 5:23-25. And again, the '783 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 7:5-8. Thus, a POSA reading the specification of the '783 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"

96. I have applied the meaning of the term "report" would have had to a POSA in view of the '783 patent's specification. Figure 1 of the '783 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:

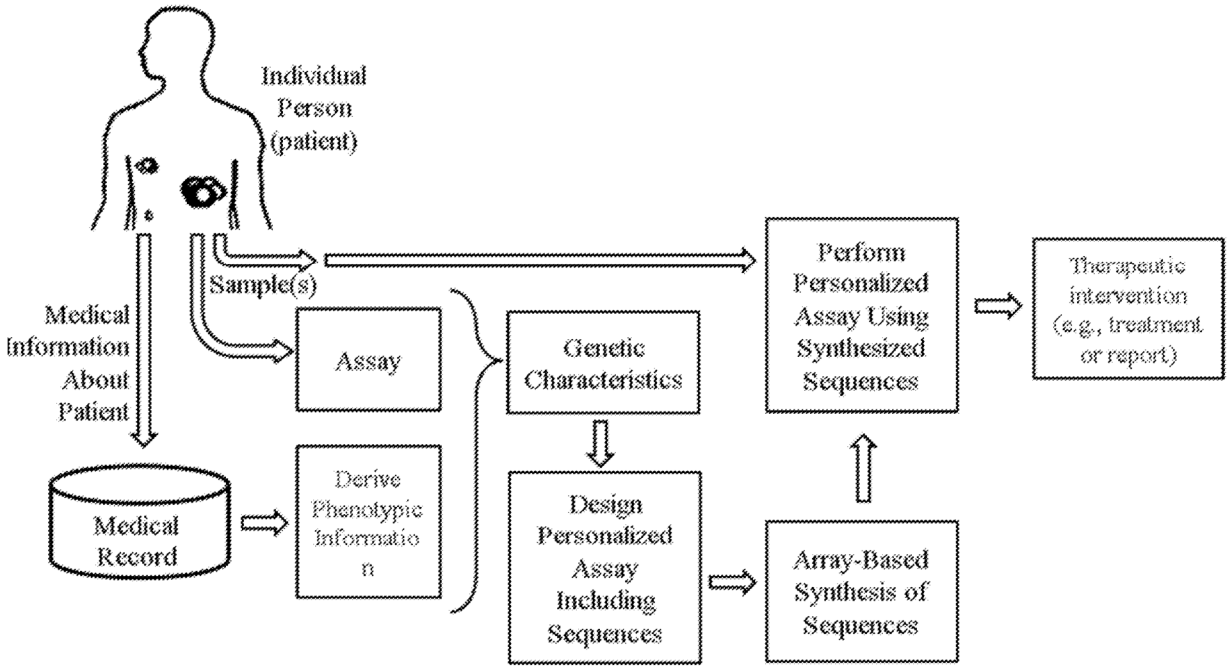


FIG. 1

97. The '783 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 5:20-23. Thus, a POSA reading the specification of the '783 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.

VII. Prior Art References of the Grounds

A. Ley (Ex. 1012)

98. 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 ’783 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 (“Personalized Cancer Testing and Monitoring”), above.

B. Forshew (Ex. 1030)

99. Forshew is an article that was published on May 30, 2012 in *Science Translational Medicine.* Ex. 1030. Forshew is therefore prior art to the ’783 patent.

C. Dawson (Ex. 1070)

100. Dawson is an article that was published in the *New England Journal of Medicine* on March 28, 2013. Ex. 1070. Dawson is therefore prior art to the ’783 patent.

VIII. The Challenged Claims Are Unpatentable

A. Ground 1: Ley Anticipates Claims 1-2, 5-7, 9, 14-15, 17-19, 21, 23-25, and 27

101. Claims 1-2, 5-7, 9, 14-15, 17-19, 21, 23-25, and 27 are anticipated by Ley.

1. Claim 1

102. Claim 1 recites “A method for personalized genetic testing, comprising” enumerated steps (a) – (d). As discussed in detail below, Ley discloses each step of such a method, and thus anticipates claim 1.

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.*

104. This method of personalized genetic testing includes each step of claim 1, as discussed below.

> **a) [*1*] obtaining a personalized probe set for a subject, which personalized probe set comprises a plurality of nucleic acid probe molecules having a plurality of nucleic acid sequences or complements thereof, [*2*]wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising a set of one or more genetic variants over other sequences in at least one biological sample of said subject, [*3*] wherein said set of genetic variants are identified in sequence data for said at least one biological sample, [*4*] further wherein said personalized probe set comprises nucleic acid probe molecules having oligonucleotide-directed genomic content comprising: (i) at least one variable portion from a result of said**

sequence data and (ii) at least one fixed portion independent of said result of said sequencing data

105. Ley discloses step 1(a) regarding "obtaining a personalized probe set for the subject" having a number of further-defined properties. I have subdivided step 1(a) into four subparts, which I address below. Because the subparts of step 1(a) relating to the properties of the personalized probe set contain overlapping features, I begin here with an overview of relevant portions of Ley's disclosure, which apply to each of the subparts of this step.

106. Ley's "Methods Summary" explains the following workflow for its method of personalized genetic testing:

> 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.

107. The first part of this workflow involves whole-genome sequencing with the Illumina Genome Analyzer and Solexa sequencing platforms 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 tumour 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 summarizes 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").

108. Following whole-genome sequencing, and "[a]fter 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." *Id.* Ley explains that this approach detected somatic mutations "in genes not previously implicated in AML pathogenesis, demonstrating the need for unbiased whole-genome approaches to discover all mutations associated with cancer pathogenesis."

Id. Ley used a several methods to analyze and filter the sequencing results, and arrived at a set of non-synonymous mutations detected in the patient's tumor sample comprising 8 SNVs and 2 insertion mutations for further analysis. *See id.* at 68-69, Table 2 (identifying mutations in the CDH24, SLC15A1, KNDC1, PTPRT, GRINL1B, GPR123, EB12, PCLKC, FLT3, and NPM1 genes).

109. Once the variants of interest, i.e., the "putative variants," were identified, Ley then describes "validation using custom PCR and capillary sequencing on the ABI 3730 platform," as well as sequencing of "validated somatic mutations" "by Roche/454 sequencing of PCR-generated amplicons made from primary genomic DNA . . . ." Ex. 1012 (Ley) at 71 ("Methods Summary"). Those skilled in the art would know that the creation of "PCR-generated amplicons made from genomic DNA" for sequencing by the Roche/454 platform necessarily involves the use of a probe set, specifically PCR primers directed to the validated somatic mutations. In Ley, the probe set would thus necessarily include PCR primers that hybridize to the 10 non-synonomous (variable) mutations identified in the patient's tumor (CDH24, SLC15A1, KNDC1, PTPRT, GRINL1B, GPR123, EB12, PCLKC, FLT3, NPM1) and the two fixed (control) mutations (BRCA2, TP53). *See* Ex. 1012 (Ley) at 69 (Fig. 3). *See* Section V.C ("Nucleic Acid Sequencing Technologies").

110. Those skilled in the art would also understand that such PCR primers necessarily constitute nucleic acid molecules that are complementary to, and as a

result hybridize to, the DNA sequences of the variant of interest. Moreover, by the nature of the PCR process, the resulting "amplicons" generated through PCR which are then sequenced on the Roche/454 platform in Ley's workflow are amplified or enriched over the other sequences in the genome which are not targeted by the set of PCR primers.

111. Subpart [1] of claimed step 1(a) is "obtaining a personalized probe set for a subject, which personalized probe set comprises a plurality of nucleic acid probe molecules having a plurality of nucleic acid sequences or complements thereof." This is disclosed by Ley.

112. As discussed above, Ley discloses "obtaining a personalized probe set for a subject," namely the set of PCR primers that are used to generate PCR amplicons for subsequent sequencing. The probes are directed to the set of genes listed in Tables 2-3 and Figure 3. Ex. 1012 (Ley) at 69-70. Because the probe set consists of PCR primers, the probe set necessarily comprises a plurality of nucleic acid probe molecules having a plurality of nucleic acid sequences or complements thereof. Specifically, the PCR primers are by their nature relatively short nucleic acid sequences that are complementary to the variant DNA sequence they are intended to amplify.

113. Subpart [2] of claimed step 1(a) is "wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences

comprising a set of one or more genetic variants over other sequences in at least one biological sample of said subject."

114. As discussed above, the nucleic acid probe molecules disclosed in Ley are PCR primers that are used to generate amplicons of the 12 variants of interest (10 variable and 2 fixed, *see* Table 3 and Figure 3) for further sequencing. Ley reports that such PCR primers are indeed used to generate PCR amplicons, which necessarily involves amplifying those sequences over other sequences in the sample. Specifically, Ley uses its PCR primers to selectively amplify the 12 variants of interest over the other sequences present in the primary tumor, skin sample, and relapse tumor sample. *See, e.g.,* Ex. 1012 (Ley) at 69 (Fig. 3).

115. Subpart [3] of claimed step 1(a) is "wherein said set of genetic variants are identified in sequence data for said at least one biological sample."

116. As discussed above, Ley identified 10 genetic mutations that are non-synonymous somatic mutations in the CDH24, SLC15A1, KNDC1, PTPRT, GRINL1B, GPR123, EB12, PCLKC, FLT3, and NPM1 genes in his initial whole-genome sequencing assay, which involved analyzing the mutations detected in both the primary tumor and skin samples and then applying a filtering process. Ex. 1012 (Ley) at 68-69. He also identified two "known" SNPs (genetic variants) in the BRCA2 and TP53 genes. Ex. 1012 (Ley) at 69 (Control amplicons containing a known heterozygous SNP in BRCA2 (encoding N372H) and a homozygous SNP in

TP53 (encoding P72R) were analysed similarly."). A POSA would understand that these two SNPs were "known" because they were previously identified in the literature from, e.g., population-level sequencing data. These two known SNPs are also "genetic variants … identified in sequence data for said at least one biological sample." Ex. 1012 (Ley), Figure 3. Accordingly, the variants ultimately included in Ley's probe set were identified in sequence data – namely whole genome sequence data – for at least one biological sample. Ley then explains that "[t]o better define the percentage of tumour cells that contained each of the discovered somatic mutations," i.e., the somatic mutations that Ley identified in the primary tumour based on the whole-genome sequencing assay, "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." *Id.* at 69. A POSA would have immediately understood that the amplification of such sequences involves increasing the relative amount of the target DNA over the "non-amplified" genomic DNA present in the samples (each of the tumor samples and the skin sample). As discussed further below, "[t]he resulting amplicons were sequenced . . . ." *Id.* An "amplicon" refers to a nucleic acid sequence that is generated in abundance using PCR, and Ley's "Methods Summary" explains that these "discovered somatic mutations" "were further analysed by Roche/454 sequencing of ***PCR-generated amplicons made from primary genomic DNA.***" *Id.* at 71. In

other words, Ley created amplicons of the variants that were identified using the sequence data, in order to further quantify the abundance of those mutations in the patient's tumor in different samples.

117. Subpart [4] of claimed step 1(a) is "further wherein said personalized probe set comprises nucleic acid probe molecules having oligonucleotide-directed genomic content comprising: (i) at least one variable portion from a result of said sequence data and (ii) at least one fixed portion independent of said result of said sequencing data"

118. 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 ("Nucleic Acid Sequencing Technologies").

119. With respect to the claimed requirement [4](i), the oligonucleotide-directed genomic content in Ley's PCR primers comprises the claimed "variable portion from a result of said sequence data." Specifically, Ley discloses including primers for the non-synonomous 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* Ex. 1012 (Ley) 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.").

120. With respect to the claimed requirement [4](ii), the oligonucleotide-directed genomic content of Ley's PCR primers also comprises "at least one fixed portion independent of said result of said sequencing data." For example, 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 inclusion of these control amplicons were based on known genetic variants, which result in known changes in the protein encoded. 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. *See also* Exs. 1073-1074 (TP53); Ex. 1076 (BRCA2). 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.

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

121. Ley discloses step 1(b). Specifically, using the PCR probes described above, Ley "amplified each mutation-containing locus from non-amplified DNA derived from the *de novo* and relapse tumour samples, and from the skin biopsy

obtained at presentation." Ex. 1012 (Ley) at 69. By amplifying the mutation-containing loci from non-amplified DNA, Ley thus discloses the amplification of such sequences over the other (non-amplified) sequences in a biological sample. Because Ley amplifies the loci of interest in (i) primary tumor; (ii) skin biopsy at presentation; and (iii) relapse tumour, Ley discloses the use of the personalized probe set in at least an additional biological sample, i.e., the relapse tumour and the skin biopsy. The PCR amplicons are used to generate a 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)." *Id.* at 69.

c) "subjecting said sequencing library to sequencing to identify a presence or absence of at least a subset of said set of genetic variants in at least said additional biological sample from said subject, wherein a sequencing footprint of said sequencing of (c) is less than a sequencing footprint of said sequence data of (a); and"

122. The sequencing library created by amplifying the relapsed tumour sample with the set of PCR primers identified the presence or absence of at least a subset (namely, all of the genetic variants included in Ley's panel) of genetic variants. 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)."). The presence or absence of the genetic variants in the multiple samples is depicted in Figure 3, which is reproduced below. Figure 3 shows the relative frequency of each given variant in each sample based on the results of the sequencing assay.

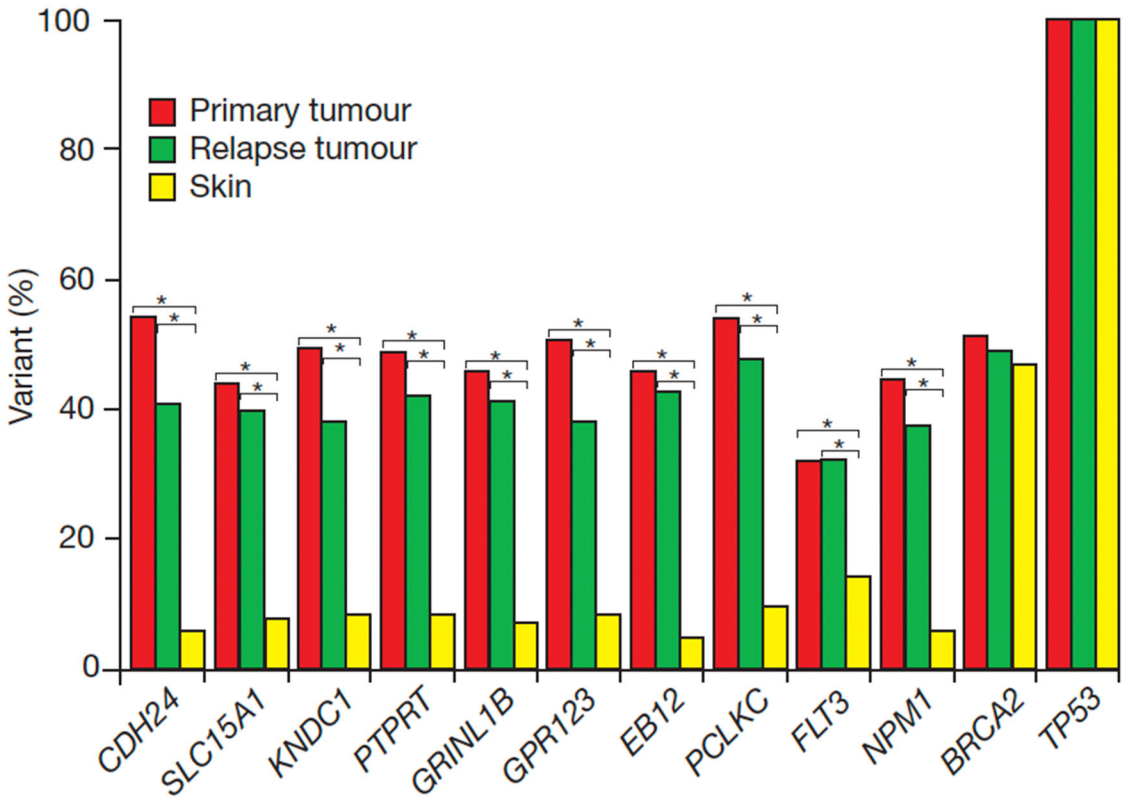


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that 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. The sequencing footprint of the sequencing performed in step ("(c)") is less than a sequencing footprint of the "sequence data of (a)." Specifically, the sequence data referenced in step (a) in Ley is derived from whole-genome sequencing. This whole-genome sequencing data necessarily has a larger sequencing footprint than the targeted sequencing performed following amplification in this step (i.e., step 1(c)). In Ley's methods, the sequencing of step (c) is performed using PCR amplicons that track the individual genes of interest identified in step (a), and thus are only small subsets of the whole genome. Accordingly, the sequencing footprint of the sequencing data of (c), which is generated by sequencing such amplicons, is necessarily less than the sequencing footprint of step (a)'s whole-genome sequence data. This is consistent with how Patent Owner reads this limitation onto the accused products. Ex. 1072 (Patent Owner's Infringement Contentions) at 4-5 (contending that limitation is met where variants "were tracked with probes of 120bp in length" while "the first sequencing step used to identify said variants (in the case of Foresight, whole genome sequencing)" has a larger footprint).

124. I have quantified the footprint of the panel of genes monitored in Ley's methods, which further confirms that "a sequencing footprint of said sequencing of (c) is less than a sequencing footprint of said sequence data of (a)." In Ley's Table 3, the panel of ten genes analyzed with Ley's targeted resequencing approach were genes from which an exon containing a genetic variant were identified via the initial

whole-genome sequencing assay. In the resequencing assay, these ten variants were targeted, to the exclusion of the other genetic material analyzed in the initial whole-genome sequencing step. 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 rather than the genome in its entirety.While Ley does not provide the primer sequences or the exact exons that were analyzed, one can deduce 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. To illustrate, I looked up the messenger RNA sequences corresponding to the coding region of these ten genes at the National Center for Biotechnology Information (NCBI), and was able to obtain the length of the coding sequence for each of the genes. GenBank records for each of these genes are also attached as exhibits 1048-1059.

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 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 bases. Thus, the sequencing footprint is far less than the ~3 billion bases sequenced in Ley's whole-genome sequencing step.

d) "generating a report identifying (i) a presence or absence of a health condition or disease of the subject in said additional biological sample based on said identified at least a subset of (c) and (ii) genetic variants which inform a therapy choice or a change in therapy for the subject."

125. Ley further discloses generating a report that shows the presence or absence of a health condition or disease of the subject in the relapsed tumor sample. The Ley publication itself is a detailed analysis of sequencing data obtained from the AML patient whose genome was subject to whole genome sequencing. Ley discloses generating a report, for example, through tabular summary of results. Specifically, Ley's Table 1 is a table summarizing various results of sequencing 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 (c) (e.g., "SNVs detected with respect to hg18 (no Y)" for "Tumor" and "Skin" samples). *Id*. at 68.

126. 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.

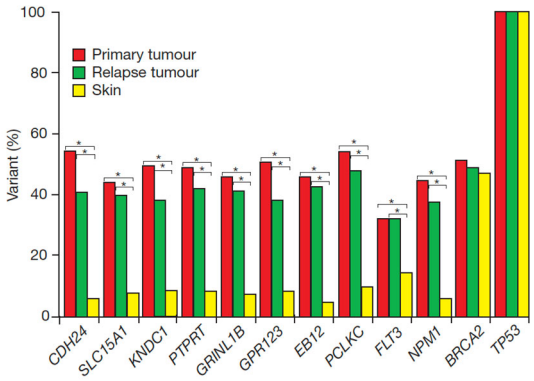


Figure 3 | Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated SNPs in the primary tumour, relapse tumour and skin specimens. The readcount data for the variant alleles in the primary tumour sample and relapse tumour sample are statistically different from that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

127. Ley's report also discloses genetic variants which inform a therapy choice or a change in therapy for the 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.

128. 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 patient’s 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.”

129. In the Supplementary Information, Ley also presents the clinical and therapeutic implication of mutations to the gene SCL29A1. "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." 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.

130. Further, one skilled in the art would know 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 (Davies 2007).

131. Accordingly, Ley anticipates claim 1.

2. Claim 2[5]

132. Claim 2 recites, "The method of claim 1, wherein said at least one biological sample comprises a plurality of biological samples."

133. Ley performs its methods on at least the primary tumor sample, the relapse tumor sample, and the skin biopsy sample. Thus, Ley analyzes a plurality of biological samples.

[5] For this and each subsequently-analyzed dependent claim, I incorporate by reference my analysis of the claim(s) from which the dependent claim depends.

134. Accordingly, Ley anticipates claim 2.

3. Claim 5

135. Claim 5 recites, "The method of claim 2, further comprising generating said sequence data of (a) from the plurality of biological samples."

136. As discussed above, Ley generates the sequence data of (a) through analysis of both the primary tumor sample and the skin biopsy sample using whole-genome sequencing. *See also* Ex. 1072 (Patent Owner contending that this claim is met in Petitioner's methods because "[t]he Foresight Diagnostics Solid Tumor Recurrence test utilizes both tumor biopsy as well as germline DNA in order to identify genetic variants for monitoring."). As discussed above, Ley's skin biopsy sample is included to show the contents of the patient's germline DNA.

137. Accordingly, Ley anticipates claim 5.

4. Claim 6

138. Claim 6 recites, "The method of claim 1, wherein said sequence data of (a) comprises sequence information for at least an exome, a panel of genes, or a whole genome."

139. As explained above with respect to claim 1, Ley's "sequence data of (a)" is sequence information from whole genome sequencing.

140. Accordingly, Ley anticipates claim 6.

5. Claim 7

141. Claim 7 recites, "The method of claim 1, wherein identifying said presence or absence of a health condition or disease is based on a frequency of a genetic variant from said set of genetic variants."

142. Ley's Table 3 shows the detection and quantification of variant genes in both primary tumor and relapsed tumor, as well as non-tumor (skin) tissue. Identifying the presence or absence of a health condition or disease (i.e., cancerous cells in a sample) is based on the frequency of one or more genetic variants from the set of variants. For example, the 10 non-synonymous mutations are detected in tumor samples but not in the skin. The "control" BRCA2 and TP53 variants are detected at their expected levels in both tumor and normal tissues.

143. Accordingly, Ley anticipates claim 7.

6. Claim 9

144. Claim 9 recites, "The method of claim 1, wherein said health condition or disease is cancer."

145. The health condition or disease addressed by Ley is acute myeloid leukemia, which is a form of cancer.

146. Accordingly, Ley anticipates claim 9.

7. Claim 14

147. Claim 14 recites, "The method of claim 1, wherein said at least one additional biological sample in (c) was obtained after said at least one biological sample in (a)."

148. The relapse tumor sample that is subjected to targeted resequencing in step 1(c) in Ley was necessarily obtained after the primary tumor sample that was subjected to whole-genome sequencing in step 1(a).

149. Accordingly, Ley anticipates claim 14.

8. Claim 15

150. Claim 15 recites, "The method of claim 1, wherein said additional biological sample is a blood sample comprising: (i) cfDNA, (ii) cfRNA, (iii) DNA or RNA from circulating tumor cells, or (iv) RNA or DNA from a tumor metastasis."

151. The additional biological sample – Ley's relapse tumor sample – necessarily contains both RNA and DNA from a tumor metastasis.

152. Accordingly, Ley anticipates claim 15.

9. Claim 17

153. Claim 17 recites, "The method of claim 1, wherein said at least one fixed portion comprises one or more known tumor variants."

154. The fixed portion of Ley comprises probes for *BRCA2* and *TP53.* These two mutations were known tumor variants in the AML genome. Ley states that "[t]he

BRCA2 variant is a known heterozygous SNP in this [AML] genome, and the *TP53* variant is a known homozygous SNP" in the AML genome. Ex. 1012 at 70 (Table 3). *See also id.* at 69 ("[N]early 100% of the *TP53* alleles were variant in all three samples, ***as expected***."). The literature also confirmed that *BRCA2* and *TP53* were known AML tumor variants. *See* Ex. 1073 (Haferlach 2008) and Ex. 1074 (Misasa 1998) (confirming knowledge of such variants). The fixed portion of Ley's panel therefore comprises two known tumor variants.

155. Accordingly, Ley anticipates claim 17.

10. Claim 18

156. Claim 18 recites, "The method of claim 1, wherein said at least one fixed portion is selected from a group consisting of one or more of: (i) cancer driver genes, (ii) genes involved in the pharmacogenomics of cancer drugs, (iii) genes involved in Mendelian immunological diseases, (iv) genes related to inherited forms of cancer, (v) genes associated with tumor escape from a targeted or immune cancer therapy, (vi) HLA typing, (vii) variants common in the population and used by B-allele methods to detect structural variation, and any combination thereof."

157. As discussed above, including with respect to claim 17, the *BRCA2* and *TP53* genes included in Ley's fixed panel are genes related to inherited forms of cancer, including AML.

158. Accordingly, Ley anticipates claim 18.

11. Claim 19

159. Claim 19 recites, "The method of claim 1, wherein said sequence data of (a) is derived from sequences of nucleic acids from said at least one biological sample, wherein said at least one biological sample comprises a tumor biopsy."

160. 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.

161. Accordingly, Ley anticipates claim 19.

12. Claim 21

162. Claim 21 recites, "The method of claim 1, wherein said set of one or more genetic variants comprise one or more members selected from the group consisting of: (i) a single nucleotide polymorphism, (ii) an insertion or deletion, (iii) a copy number variation, and (iv) structural variation."

163. The genetic variants analyzed in Ley's panel include at least SNPs and insertion or deletions. *See, e.g.*, Ex. 1012 (Ley) at 69 (Table 2). A POSA would understand that a missense mutation, such as those listed in Ley's Table 2, is a SNP that results in a non-synonymous mutation, meaning that it would change the protein encoded by the SNP-containing gene and potentially altering its function.

164. Accordingly, Ley anticipates claim 21.

13. Claim 23

165. Claim 23 recites, "The method of claim 1, wherein said set of genetic variants identified in sequence data are determined with respect to a reference."

166. Ley discloses comparing the sequence data from the primary tumor sample to a sequenced skin sample, which constitutes a reference. *See* Ex. 1012 (Ley) at 68-69 ("Identification of somatic mutations in coding sequences"). Ley also describes comparing the sequence data to other references, such as previously described human genomes. *Id.*; *see also id.* at 67 ("Coverage depth of the tumour and skin genomes") ("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).").

167. Accordingly, Ley anticipates claim 23.

14. Claim 24

168. Claim 24 recites, "The method of claim 1, further comprising, prior to (a), sequencing said least one biological sample to generate said sequencing data."

169. As explained above, Ley discloses performing whole-genome sequencing on the primary tumor sample prior to step 1(a) in order to generate said sequencing data.

170. Accordingly, Ley anticipates claim 24.

15. Claim 25

171. Claim 25 recites, "The method of claim 1, further comprising generating said personalized probe set for said subject based on said set of genetic variants identified in said sequence data for said at least one biological sample."

172. As discussed above, the personalized probe set disclosed in Ley is based on the sequencing data identified in the primary tumor using whole-genome 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* Ex. 1012 (Ley) 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.”).

173. Accordingly, Ley anticipates claim 25.

16. Claim 27

174. Claim 27 recites, “The method of claim 1, wherein (c) further comprises identifying levels of gene expression, sequencing read counts, or read depth.”

175. Ley’s Tables 2 and 3 disclose identifying sequencing read counts for each of the variants. Ex. 1012 (Ley) at 69 (Table 2) (annotated below); *id.* at 70 (Table 3) annotated below).

Table 2 | Non-synonymous somatic mutations detected in the AML sample

Gene	Consequence	Type	Solexa tumour reads WT:variant	Solexa skin reads WT:variant	Conservation score of mutant base	Mutations in other AML cases*
CDH24	Y590X	Nonsense	9:9	16:0	0.998	0/187
SLC15A1	W77X	Nonsense	15:12	19:0	1.000	0/187
KNDC1	L799F	Missense	7:8	20:0	NA	0/187
PTPRT	P1235L	Missense	9:13	16:0	1.000	0/187
GRINL1B	R176H	Missense	15:10	14:0	NA	0/187
GPR123	T38I	Missense	11:11	13:0	NA	0/187
EBI2	A338V	Missense	7:12	18:2	1.000	0/187
PCLKC	P1004L	Missense	19:9	15:1	0.98	0/187
FLT3	ITD	Indel	18:12	8:0	NA	51/185
NPM1	CATG ins	Indel	36:6	33:0	NA	43/180

Ins, insertion; WT, wild type.
* Patient cohort defined in ref. 23.

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

		Primary AML (100% blasts)			Skin			Relapse (78% blasts)		
Gene	Consequence	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
KNDC1	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.

176. Ley also discloses further identifying levels of gene expression in both the primary and relapse tumor samples. Specifically, Ley states “seven of the ten

genes containing somatic mutations were detectably expressed in the tumour sample. *FLT3* and *NPM1* messenger RNAs were highly expressed in this tumour sample, as they are in virtually all AML samples. We detected mRNA from the *CDH24*, *SIC15A1* and *EBI2* genes on the Affymetrix expression array, whereas expression of *GRINL1B* and *PCLKC* were detected by PCR with reverse transcription (RT-PCR; data not shown). Expression of *KNDC1*, *PTPRT* and *GPR123* was not detected by either approach, but we cannot rule out expression of these genes in a small subset of tumour cells (for example, leukaemia-initiating cells)." Ex. 1012 (Ley) at 70. A POSA understood that both the detection of mRNA expression and RT-PCR were techniques used to monitor gene expression.

177. Accordingly, Ley anticipates claim 27.

B. Ground 2: Obviousness Over Ley in view of a POSA's Knowledge

178. All Challenged Claims are obvious over Ley in view of the knowledge of a POSA for the reasons set forth below. I incorporate by reference my anticipation analysis of Ley as to the claims addressed in Ground 1. A POSA would also have found the Challenged Claims obvious for the additional reasons set forth below.

1. Claim 1

179. As explained above, the disclosures of Ley anticipate claim 1. In light of a POSA's knowledge as of the May 2016 priority date of the Challenged Claims, at least the following additional modifications to Ley would have been obvious.

180. First, at an overall level, a POSA would have found it obvious to apply Ley's approach – whole-genome sequencing followed by targeted serial monitoring of tumor-specific mutations identified in such whole-genome sequencing – to a variety of cancers other than the specific example of AML discussed in Ley. Indeed, such approaches were widely used before the priority date. *See* Section V.D ("Personalized Cancer Testing and Monitoring"); Ex. 1031 (Meyerson); Ex. 1002 (Leary); Ex. 1015 (McBride).

181. In applying Ley's methods to AML and other cancers, the following additional modifications would have been obvious:

182. **Step 1(a) – "fixed portion"**: 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.

183. 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.").

184. 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.

185. 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.

186. 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.

187. 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.

188. **Step 1(d) – "generating a report identifying (i) a presence or absence of a health condition or disease of the subject in said additional biological sample based on said identified at least a subset of (c) and (ii) genetic variants which inform a therapy choice or a change in therapy for the subject.":** A POSA would have also found it obvious, based on the results of the genetic testing methods disclosed and rendered obvious by Ley, to generate a report identifying a presence or absence of a health condition based on the results of assaying a particular sample, and to identify genetic variants that inform a therapy choice or change in therapy for the subject. As an initial matter, although Ley expressly discloses testing biological samples at the onset of cancer and in a relapsed tumor, one skilled in the

art would have found it obvious to perform Ley's assays at multiple timepoints throughout the progression of the cancer in order to understand the course of the disease. Doing so would have involved only performing the tests on samples taken at different time points, and was successfully performed in multiple other studies that followed Ley and preceded the priority date. *See* Section V.D ("Personalized Cancer Testing and Monitoring").

189. As explained in the anticipation ground above, 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."

190. 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." 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.

191. Further, one skilled in the art would know that the characteristics of the disease and the patient's positive mutation 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 (Davies 2007).

192. 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 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,

and 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.

193. For these reasons, in addition to anticipating claim 1, Ley in view of the knowledge of a POSA as of the priority date would have also rendered obvious additional variations of Ley's methods.

2. Claim 2

194. Claim 2 recites, "The method of claim 1, wherein said at least one biological sample comprises a plurality of biological samples."

195. As discussed above, Ley discloses performing targeted resequencing using the panel of variable and fixed probes on primary and relapsed tumor samples. 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 additional time points, resulting in the collection of more than two biological sample (i.e., a plurality) following the initial sample for analysis and sequencing using the Ley's personalized probe sets. A POSA would have reasonably expected

success in doing 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 Ley, 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).

196. Accordingly, Ley in view of a POSA's knowledge renders claim 2 obvious.

3. Claim 3

197. Claim 3 recites, "The method of claim 2, wherein said plurality of biological samples comprises a tissue sample and a plasma sample."

198. I incorporate by reference my discussion of claim 2. Ley discloses collecting samples from tissue samples, specifically a tumour sample. *See, e.g.,* Ex. 1012 (Ley) at 69. However, those skilled in the art would have been familiar with the release of tumor DNA into the bloodstream, resulting in cell-free tumor DNA circulating in the blood. *See* Section V.E ("Related Applications of Next Generation Sequencing"). A POSA would have been motivated to use plasma samples -- where

such cell-free DNA is present in solution – in addition to or instead of tissue samples like tumor samples to ease sample collection. A POSA would have reasonably expected success in applying Ley's methods to biological samples derived from plasma in light of the successful reports of others, like Leary (Ex. 1002) and McBride (Ex. 1015) doing so in the serial monitoring of tumor DNA in the blood plasma of patients with solid tumors.

199. Accordingly, Ley in view of a POSA's knowledge renders claim 3 obvious.

4. Claim 4

200. Claim 4 recites, "The method of claim 2, wherein the plurality of biological samples comprise: (i) cell-free deoxyribonucleic acid molecules (cfDNA), (ii) cell-free ribonucleic acid molecules (cfRNA), (iii) DNA or RNA from circulating tumor cells, or (iv) RNA or DNA from a tumor metastasis."

201. I incorporate by reference my discussions of claims 2 and 3 above. As I note above, it would have been obvious to specifically monitor cell free DNA in a plasma sample in Ley's methods. Moreover, in metastasized cancer, where the cancer has spread to multiple parts of the body, a POSA would have found it obvious to take biopsies from tumors located in different locations. As a result, it also would have been obvious in view of Ley to take biological samples that would thus necessarily include RNA or DNA from a tumor metastasis.

202. Accordingly, Ley in view of a POSA's knowledge renders claim 4 obvious.

5. Claim 8

203. Claim 8 recites, "The method of claim 1, wherein said additional biological sample of (c) is a blood sample."

204. Ley discloses collecting samples from tissue samples, specifically a tumour sample. *See, e.g.,* Ex. 1012 (Ley) at 69. However, those skilled in the art would have been familiar with the release of tumor DNA into the bloodstream, resulting in cell-free tumor DNA circulating in the blood. *See* Section V.E ("Related Applications of Next Generation Sequencing"). A POSA would have been motivated to use plasma samples -- where such cell-free DNA is present in solution – in addition to or instead of tissue samples like tumor samples to ease sample collection. A POSA would recognize that a plasma sample is a type of blood sample. A POSA would have reasonably expected success in applying Ley's methods to biological samples derived from plasma in light of the successful reports of others, like Leary (Ex. 1002) and McBride (Ex. 1015) doing so in the serial monitoring of tumor DNA in the blood plasma of patients with solid tumors.

205. Accordingly, Ley in view of a POSA's knowledge renders claim 8 obvious.

6. Claim 13

206. Claim 13 recites, "The method of claim 1, further comprising repeating (b) and (c) one or more times and generating one or more additional reports."

207. As discussed above, Ley discloses performing targeted resequencing using the panel of variable and fixed probes on primary and relapsed tumor samples. 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 additional time points, resulting in the collection of more than two biological sample following the initial sample for analysis and sequencing using Ley's personalized probe sets. A POSA would have reasonably expected success in doing 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). Moreover, those studies disclosed generating reports of the results at each time point.

208. Accordingly, Ley in view of a POSA's knowledge renders claim 13 obvious.

7. Claim 14

209. Claim 14 recites, "The method of claim 1, wherein said at least one additional biological sample in (c) was obtained after said at least one biological sample in (a)."

210. As explained above, Ley anticipates claim 14. Moreover, when performing the methods rendered obvious by Ley to serially monitor a patient's cancer at 3 or more time points, the additional biological samples will necessarily be obtained after the initial biological sample, in order to monitor the cancers progression.

211. Accordingly, Ley in view of a POSA's knowledge renders claim 14 obvious.

8. Claim 15

212. Claim 15 recites, "The method of claim 1, wherein said additional biological sample is a blood sample comprising: (i) cfDNA, (ii) cfRNA, (iii) DNA or RNA from circulating tumor cells, or (iv) RNA or DNA from a tumor metastasis."

213. I incorporate by reference my discussions of claims 2 and 3 above. As I note above, it would have been obvious to specifically monitor cell free DNA in a plasma sample in Ley's methods. Moreover, the presence of components like

cfDNA, cfRNA, and circulating tumor nucleic acids is an inherent result of cancer biology, and such substances will necessarily be present in the blood and blood plasma of cancer patients.

214. Additionally, in metastasized cancer, where the cancer has spread to multiple parts of the body, a POSA would have found it obvious to take biopsies from tumors located in different locations. Accordingly, it also would have been obvious in view of Ley to take biological samples that would thus necessarily include RNA or DNA from a tumor metastasis.

215. Accordingly, Ley in view of a POSA's knowledge renders claim 15 obvious.

9. Claim 16

216. Claim 16 recites, "The method of claim 1, wherein said additional biological sample and one or more subsequent biological samples are each blood samples comprising: (i) cfDNA, (ii) cfRNA, (iii) DNA or RNA from circulating tumor cells, or (iv) RNA or DNA from a tumor metastasis."

217. I incorporate by reference my discussions of claims 2, 3, and 15 above. As I note above, it would have been obvious to specifically monitor cell free DNA in a plasma sample in Ley's methods. Moreover, the presence of components like cfDNA, cfRNA, and circulating tumor nucleic acids is an inherent result of cancer

biology, and such substances will necessarily be present in the blood and blood plasma of cancer patients.

218. Additionally, in metastasized cancer, where the cancer has spread to multiple parts of the body, a POSA would have found it obvious to take biopsies from tumors located in different locations. Accordingly, it also would have been obvious in view of Ley to take biological samples that would thus necessarily include RNA or DNA from a tumor metastasis.

219. Accordingly, Ley in view of a POSA's knowledge renders claim 16 obvious.

10. Claim 18

220. Claim 18 recites, "The method of claim 1, wherein said at least one fixed portion is selected from a group consisting of one or more of: (i) cancer driver genes, (ii) genes involved in the pharmacogenomics of cancer drugs, (iii) genes involved in Mendelian immunological diseases, (iv) genes related to inherited forms of cancer, (v) genes associated with tumor escape from a targeted or immune cancer therapy, (vi) HLA typing, (vii) variants common in the population and used by B-allele methods to detect structural variation, and any combination thereof."

221. As explained above in my obviousness analysis of element 1(a)'s "fixed portion" requirement, which I incorporate by reference here, 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.

222. Accordingly, Ley in view of a POSA's knowledge renders claim 18 obvious.

11. Claim 22

223. Claim 22 recites, "The method of claim 1, wherein said personalized probe set comprises hybrid capture probes."

224. Ley discloses the use of PCR primers as probes for the amplification and/or enrichment of target DNA over normal DNA in the sample to prepare its sequencing library. However, by the 2016 priority date of the '783 patent, a POSA would have also been familiar with the use of hybrid capture techniques for similar purposes. As explained in Section V.C.3, above, PCR and hybrid capture 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"). Accordingly, a POSA would have found it obvious to use create hybrid capture probes targeting the same sequences as Ley because doing so would have involved only the predictable usage of techniques known in the art to be interchangeably useful for the same purpose. Indeed, 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 2011) at 379. Thus, a POSA would have readily been able to obtain "out of the box" hybrid capture reagents for a "custom enrichment region" including a panel of probes directed to the genetic variations one would monitor in Ley's methods and the variations thereof discussed in this section.

225. The '783 patent's specification confirms that those skilled in the art considered these techniques to be interchangeable as of the priority date. In defining the terms "probe," "nucleic acid probe," "capture probe," and "bait," the '783 patent explains:

> The terms "probe," "nucleic acid probe," "capture probe," "bait," as used herein, generally refer to a nucleic acid molecule comprising a single-stranded portion capable of hybridizing to a complementary nucleic acid sequence. A probe can be used for detection or enrichment of nucleic acid molecules. A probe can be target-specific such that a region of interest may be pulled-down, isolated, enriched, amplified, or labeled. A probe can be used for targeted sequencing. A probe may hybridize to a targeted sequence when attached to a solid substrate or when in-solution, e.g. as for hybrid capture. A probe may be included in a set, or plurality, of probes. A probe set can comprise probes that

> overlap within a specific genomic region such that they are tiled or staggered. A probe set can include probes to a genomic region or a panel comprising multiple genomic regions. Probes can be amplification based or capture hybridization-based. Non-limiting examples of probes include molecular inversion probes, amplification probes, biotinylated affinity probes, or any probe comprising a detectable label.

Ex. 1001 at 29:4-22.

226. Accordingly, Ley in view of a POSA's knowledge renders claim 22 obvious.

12. Claim 26

227. Claim 26 recites, "The method of claim 1, wherein (b) further comprises using said personalized probe set to selectively enrich or amplify sequences comprising said set of genetic variants over other sequences in a biological sample from another subject to generate said sequencing library."

228. As explained above, applications of similar sequencing technologies were known to those skilled in the art to be applicable to the monitoring of cell-free DNA in various contexts, including organ transplant rejection and prenatal diagnostics. *See* Section V.E ("Related Applications of Next Generation Sequencing"). In those scenarios, when there are foreign cells present in the body – whether from a fetus or from a donor organ – those cells release into the blood stream

cell free DNA with its own signature. In the context of a cancer patient that was also pregnant or the recipient of an organ transplant, a POSA would have found it obvious to selectively enrich or amplify sequences comprising the genetic variants of the patient's cancer over other sequences that may be present in the sample from another subject – such as circulating fetal DNA or organ donor-derived DNA – in order to generate the sequencing library. A POSA would have been motivated to select sequences unique to the tumor, as opposed to the foreign DNA from a fetus or donor organ, in order to precisely monitor the cancer over other medical issues of concern.

229. Accordingly, Ley in view of the knowledge of a POSA renders claim 26 obvious.

13. Claim 28

230. Claim 28 recites, "The method of claim 1, wherein said at least one additional biological sample comprises a plurality of additional samples obtained at different time points."

231. As discussed above, Ley discloses performing targeted resequencing using the panel of variable and fixed probes on primary and relapsed tumor samples. 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 additional time points, resulting in the collection of more than two biological sample (i.e., a plurality) following the initial sample for analysis and sequencing

using Ley's personalized probe sets. A POSA would have reasonably expected success in doing 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). Moreover, those studies disclosed generating reports of the results at each time point.

232. Accordingly, Ley in view of a POSA's knowledge renders claim 28 obvious.

14. Claim 29

233. Claim 29 recites, "The method of claim 1, wherein the report of (d) further comprises reporting biomedical information of said subject that is predictive, prognostic, or diagnostic of one or more biomedical features selected from the group consisting of a disease state, efficacy of a drug therapy, prediction of optimal drug dosage, a recommendation of one or more therapies, and a recommendation of a course of treatment of a disease."

234. A POSA would have also found it obvious to, based on the results of the genetic testing in methods disclosed and rendered obvious by Ley, generate a report identifying a presence or absence of a health condition based on the results of assaying a particular sample, and to identify genetic variants that inform a therapy choice or change in therapy for the subject. As an initial matter, although Ley expressly discloses testing biological samples at the onset of cancer and in a relapsed tumor, one skilled in the art would have found it obvious to perform Ley's assays at multiple timepoints throughout the progression of the cancer in order to understand the course of the disease. Doing so would have involved only performing the tests at samples taken at different time points, and was successfully performed in multiple other studies that followed Ley and preceded the priority date. *See* Section V.D ("Personalized Cancer Testing and Monitoring").

235. As explained above in Section V.F ("Providing Actionable Information to Healthcare Providers"), an important aspect of reporting the results of cancer-monitoring assays is providing actionable information to healthcare providers. Thus, 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, and it would have been obvious to report similar information at each time point for which a sample was collected and analyzed 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. Such a report would therefore necessarily provide information relating to at least the claimed "a recommendation of one or more therapies, and a recommendation of a course of treatment of a disease."

236. Accordingly, Ley in view of a POSA's knowledge renders claim 29 obvious.

C. Ground 3: Obviousness over Forshew in view of Dawson

1. Claim 1

237. Claim 1 recites "A method for personalized genetic testing, comprising" a number of additional steps. Forshew discloses such a method by teaching to analyze multiple nucleic acid samples isolated from an individual by "demonstrat[ing] use of TAm-Seq [*i.e.*, Forshew's amplification and deep sequencing of nucleic acid technique]" "to monitor tumor dynamics, and tracked 10 concomitant mutations in plasma of a metastatic breast cancer patient over 16

months." Ex. 1030 at Abstract. This method, Forshew noted, "could facilitate analysis of circulating DNA as a noninvasive 'liquid biopsy' for personalized cancer genomics." *Id.*

238. Moreover, as discussed further below, Forshew in view of Dawson discloses the claimed method steps, and hence renders the claim obvious.

> **a) [*1*] obtaining a personalized probe set for a subject, which personalized probe set comprises a plurality of nucleic acid probe molecules having a plurality of nucleic acid sequences or complements thereof, [*2*]wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising a set of one or more genetic variants over other sequences in at least one biological sample of said subject, [*3*] wherein said set of genetic variants are identified in sequence data for said at least one biological sample, [*4*] further wherein said personalized probe set comprises nucleic acid probe molecules having oligonucleotide-directed genomic content comprising: (i) at least one variable portion from a result of said sequence data and (ii) at least one fixed portion independent of said result of said sequencing data**

239. Forshew discloses step 1(a) regarding "obtaining a personalized probe set for the subject" having a number of further-defined properties. I have subdivided step 1(a) into four subparts, which I address below. Because the subparts of step 1(a) relating to the properties of the personalized probe set contain overlapping features,

I begin here with an overview of relevant portions of Forshew's disclosure, which apply to each of the subparts of this step.

240. Forshew describes a method of serial monitoring of a patient's cancer using a technique it refers to as "Tam-Seq." Forshew explains that "TAm-Seq can be flexibly adapted to sequence different genomic regions by designing primers to amplify regions of interest. We used this capability to study dynamics of multiple mutations in parallel." Ex. 1030 (Forshew) at 8. In one exemplary usage of its techniques, Forshew explained that "[w]hole-genome sequencing of tumor material was used to identify tumor mutations in a patient with metastatic breast cancer undergoing two phases of chemotherapy. Ten mutations were selected, and short amplicons (<120 bp) were designed to cover the mutation loci (table S7). Serial plasma samples were collected over the course of 497 days, both before and after treatment. We performed TAm-Seq in duplicate . . . and tracked dynamics of all mutations in parallel (Fig. 4D)." *Id.* at 9.

241. Forshew discloses producing a plurality of PCR primers (also referred to as PCR probes), wherein the plurality of PCR probes hybridize to a plurality of genetic variants in a sample to amplify or enrich them over other sequences in the sample. For example, Forshew "designed primers to generate amplicons that tile[d] regions of interest in short segments of 150 to 200 bases. . . ." Ex. 1030 at 2-3. Forshew designed primers using Primer3, which is a computer program used for

designing PCR primers (also referred to as PCR probes). Ex. 1032 at 2 ("Primer design for Tam-Seq"); Primer3, *available at* http://primer3.org (last accessed January 26, 2023); *see also* Ex. 1034 (Untergasser *et al.*, Primer3—new capabilities and interfaces, Nucleic Acids Res., 2012, 40(15)).

242. Forshew further used the COSMIC database, which is a database of human cancer sequences, to select polymorphisms to which the PCR probes (primers) would hybridize. Ex. 1030 at 3; COSMIC, *available at* https://cancer.sanger.ac.uk/cosmic (last accessed January 26, 2023); *see also* Ex. 1035 (Forbes *et al.*, COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer, Nucleic Acids Res, 2011, 39, D945-950). For example, I used the UCSC In Silico PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr) to confirm that the first pair of primer sequences in Table S1 of Forshew *et al.*'s Supplementary Material targets exon 10 of the PIKC3A gene on chromosome 3 cataloged in the COSMIC database. I confirmed, using version 4 of the COSMIC update dated November 12, 2004, that mutations in PIKC3A (second to last in the gene listing in the excerpt pasted below) was reported as of that date at the latest.

COSMIC Update

The COSMIC team are proud to include somatic mutation data for FGFR2, FGFR3, FLT3, MET, PDGFRA and PIK3CA on the COSMIC web site.

Data

The number of genes with data in COSMIC has more than doubled in this release of the database. The additional data represents a set of genes that have a lower, but nevertheless important, mutation frequency in human cancer as a whole. In specific malignancies genes such as FLT3 do have a significant role as can be seen from the data collected in COSMIC.

Gene	Number of analysed samples	Number of samples with mutations
BRAF	5158	736
ERBB2	714	8
FGFR2	30	2
FGFR3	1735	481
FLT3	7610	1499
HRAS	11876	477
KRAS2	35716	8302
MET	1081	59
NRAS	13884	1132
PDGFRA	146	25
PIK3CA	396	89
TOTAL	78346	12810

Number of unique mutations 307

Number of curated papers 976

243. Forshew also discloses the design and use of a set of probes in the form of PCR primers that hybridize to and amplify a plurality of variants (also known in the art as polymorphisms) in sequences encoding genes associated with cancer. For example, Forshew designed the PCR primers to hybridize to selected coding sequences in genes that include mutations with known associations with cancers. Ex. 1030 at 3. Forshew states that "[w]e designed a set of primer pairs to amplify 5995 bases of genomic sequence covering coding regions (exons and exon junctions) of *TP53* and *PTEN*, and selected regions of *EGFR*, *BRAF*, *KRAS*, and *PIK3CA* (table S1) by overlapping short amplicons (Fig. 1A)." The genes selected by Forshew, namely *TP53, PTEN, EGFR, BRAF, KRAS* and *PIK3CA*, were genes with known biomedically interpretable variants. Indeed, Forshew notes that "[t]he sequenced regions cover mutations that account for 38% of all point mutations in the COSMIC database (v55) (32)." Ex. 1030 at 3.

244. Forshew further discloses that the plurality of polymorphisms are based on or extracted from the COSMIC database, which, as discussed above, is a database that catalogued polymorphisms (*i.e.*, mutations result in different genetic variants) present in and associated with cancers observed in a human population. Those polymorphisms were obtained, identified, and catalogued by observation in the wide variety of reports in the scientific literature concerning cancer mutations. Ex. 1035 at 1 ("COSMIC combines cancer mutation data from the scientific literature, with the output from the Cancer Genome Project (CGP) at the Sanger Institute UK.").

245. In addition, Forshew also discloses generating such PCR probes based on whole-genome sequencing data. Ex. 1030 (Forshew) at 9 ("Whole-genome sequencing of tumor material was used to identify tumor mutations in a patient with metastatic breast cancer. . . .").

246. Accordingly, Forshew discloses producing a set of PCR probes (primers) that hybridize to a plurality of genetic variants based on multiple criteria, including the use of known genetic variants as well as through the use of an initial sequencing step to determine personalized variants specific to that patient's tumor.

247. Subpart [1] of claimed step 1(a) is "obtaining a personalized probe set for a subject, which personalized probe set comprises a plurality of nucleic acid probe molecules having a plurality of nucleic acid sequences or complements thereof."

248. Forshew discloses obtaining a plurality of personalized PCR probes for its method, which are nucleic acid molecules that have complementary sequences to those identified in the patient. In one example, Forshew performed "[w]hole-genome sequencing of tumor material" "to identify tumor mutations in a patient with metastatic breast cancer." Ex. 1030 (Forshew) at 9. From this sequence data, "[t]en mutations were selected, and short amplicons (<120 bp) were designed to cover the mutation loci." *Id.* These amplicons were tracked using the "TAm-Seq" process disclosed in Forshew, which as explained above is disclosed to require the creation of PCR probes to the genetic loci in order to generate the amplicons for sequencing. Thus, Forshew discloses or at a minimum renders obvious this requirement of step 1(a).

249. Subpart [2] of claimed step 1(a) is "wherein said plurality of nucleic acid probe molecules are configured to selectively enrich or amplify sequences comprising a set of one or more genetic variants over other sequences in at least one biological sample of said subject."

250. The PCR probes used in Forshew's TAm-Seq process are configured to amplify the sequences containing the genetic variants of interest over other sequences present in the sample. As Forshew explains, "[u]sing standard amplification primers tailored to the mutation loci, we [] used TAm-Seq to monitor the dynamics of 10 mutations in plasma DNA of a single patient with metastatic

breast cancer, using minute amounts of input DNA." Ex. 1030 (Forshew) at 10. The use of such "amplification primers" to generate amplicons necessarily involves amplifying the mutated sequences, i.e., the one or more genetic variants present in the plasma samples, over other sequences present in the sample.

251. Subpart [3] of claimed step 1(a) is "wherein said set of genetic variants are identified in sequence data for said at least one biological sample."

252. The genetic variants for which Forshew designed its TAm-Seq primers were identified in sequence data for at least one biological sample. Specifically, "[w]hole-genome sequencing of tumor material was used to identify tumor mutations in a patient," "[t]en mutations were selected," "[s]erial plasma samples were collected . . . both before and after treatment," and the TAm-Seq process was "performed . . . in duplicate," which "tracked dynamics of all mutations in parallel." Ex. 1030 (Forshew) at 9. Accordingly, the ten mutations that are tracked using TAm-Seq are variants that were identified in the whole-genome sequencing data for at least one biological sample.

253. Subpart [4] of claimed step 1(a) is "further wherein said personalized probe set comprises nucleic acid probe molecules having oligonucleotide-directed genomic content comprising: (i) at least one variable portion from a result of said sequence data and (ii) at least one fixed portion independent of said result of said sequencing data."

254. A POSA would understand that Forshew's PCR primers (probes) are nucleic acid molecules having oligonucleotide-directed genomic content, i.e., sequences that are designed to bind or hybridize to nucleic acids in a sample having a complementary sequence.

255. Forshew expressly discloses the use of a variable portion from a result of the whole-genome sequencing data in the context of its breast cancer patient. As discussed above, in the breast cancer patient whose cancer was tracked using TAm-Seq following an initial whole-genome sequencing step, the set of 10 nucleic acid probe molecules used to generate the amplicons for sequencing constitute a variable portion from a result of said sequencing data. Ex. 1030 (Forshew) at 9 ("[w]hole-genome sequencing of tumor material was used to identify tumor mutations in a patient," and "[t]en mutations were selected.").

256. In the example studying the breast cancer patient, Forshew does not expressly disclose including PCR primers (probes) for sequences or markers that constitute a fixed portion that is selected independent of the result of said sequencing data. However, it would have been obvious to include such probes in Forshew's panel for a particular patient as well. Indeed, Forshew discloses including such probes in other examples. For example, Forshew notes that in the context of ovarian cancers "mutations in the tumor suppressor gene *TP53* have been observed in 97% of cases (20, 21), but these are located throughout the gene and difficult to assay."

Ex. 1030 (Forshew) at 1. Forshew developed "[a] cost-effective method that could detect and measure allele frequency (AF) of *TP53* mutations in plasma" which may be "highly applicable as a biomarker." *Id.* Forshew disclosed including in its set of probes or primers "primer pairs to amplify 5995 basis of genomic sequence covering coding regions (exons and exon junctions) of *TP53*," among multiple other genes. *Id.* at 3. Thus, Forshew disclosed including probes for variants in *TP53* independent of any prior sequencing data in its other assays. Including such additional probes in addition to sequence-based probes in a single panel would require nothing more than the predictable use of techniques used in other examples of the very same reference.

257. Moreover, Dawson, on which Dr. Forshew is also an author, explains that *TP53* is also "known to be recurrently mutated in breast cancer." Ex. 1070 (Dawson) at 1207. Dawson further describes applying the TAm-Seq technique to detect variants of *TP53* in breast cancer patients. *Id.* ("The monitoring of circulating tumor DNA levels requires the identification of somatic alterations in individual patients. Future developments will reduce the cost of whole-genome paired-end sequencing and targeted sequencing can be readily expanded to include other genes, in addition to *PIK3CA* and *TP53*, known to be recurrently mutated in breast cancer."). *See also, e.g., id.* at Fig. 2.

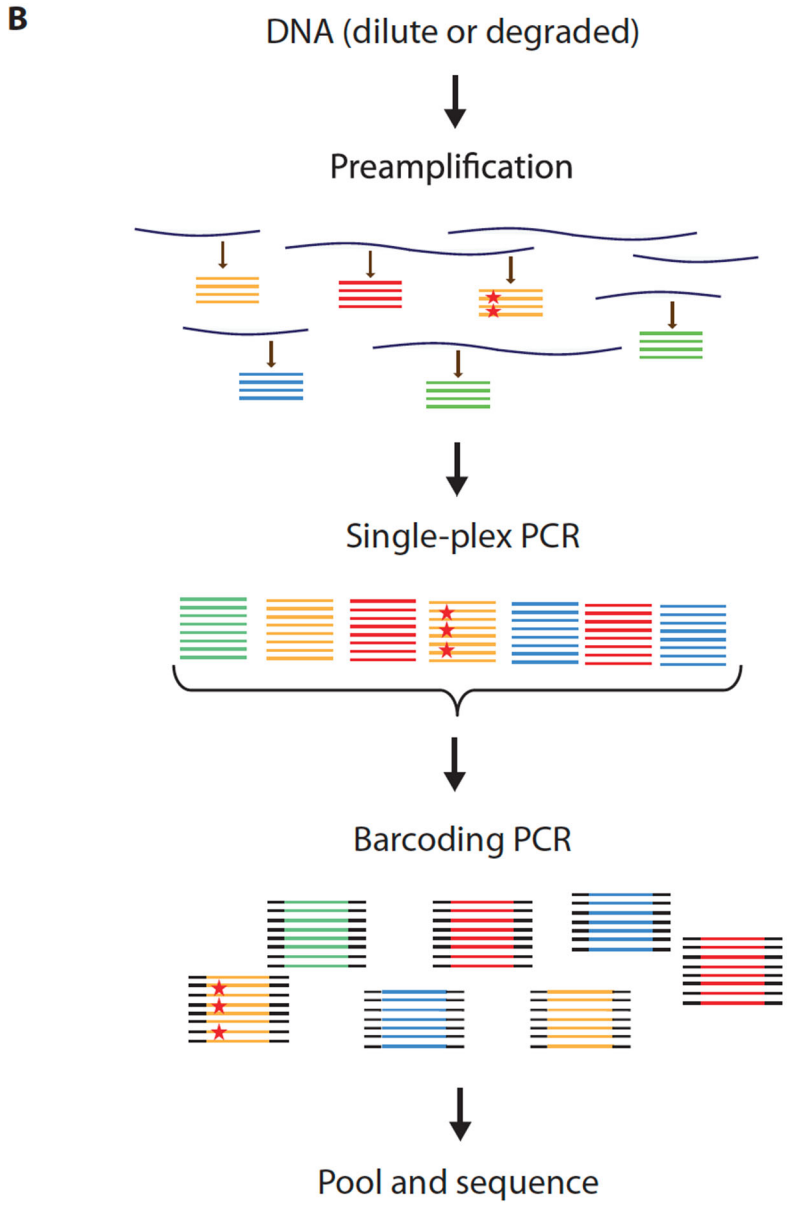
258. Accordingly, a POSA would have found it obvious to include one or more "fixed" probes in Forshew's breast cancer panel independent of any

sequencing data in order to detect and amplify variants of *TP53* in the patient. In light of Forshew's disclosure, the design of such probes was known, and would have involved only routine skill in the art and the application of the exact techniques disclosed in the article, and which were applied to *TP53* variants in breast cancer in Dawson. Accordingly, a POSA would have had a reasonable expectation of success in monitoring *TP53* using probes designed independent of Forshew's whole-genome sequencing data given that it would simply involve the use of the techniques already disclosed in a related context.

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

259. Forshew's TAm-Seq process involves using the personalized probe set (i.e., the set of PCR primers) to selectively enrich or amplify the selected genetic variants over other sequences in at least an additional biological sample from the subject. Specifically, Forshew discloses using its methods to "track[] dynamics of all mutations" in "[s]erial plasma samples [] collected over the course of 497 days, both before and after treatment." Ex. 1030 (Forshew) at 9. In Forshew's workflow, the primers are used to amplify the sequences of interest, before they are barcoded

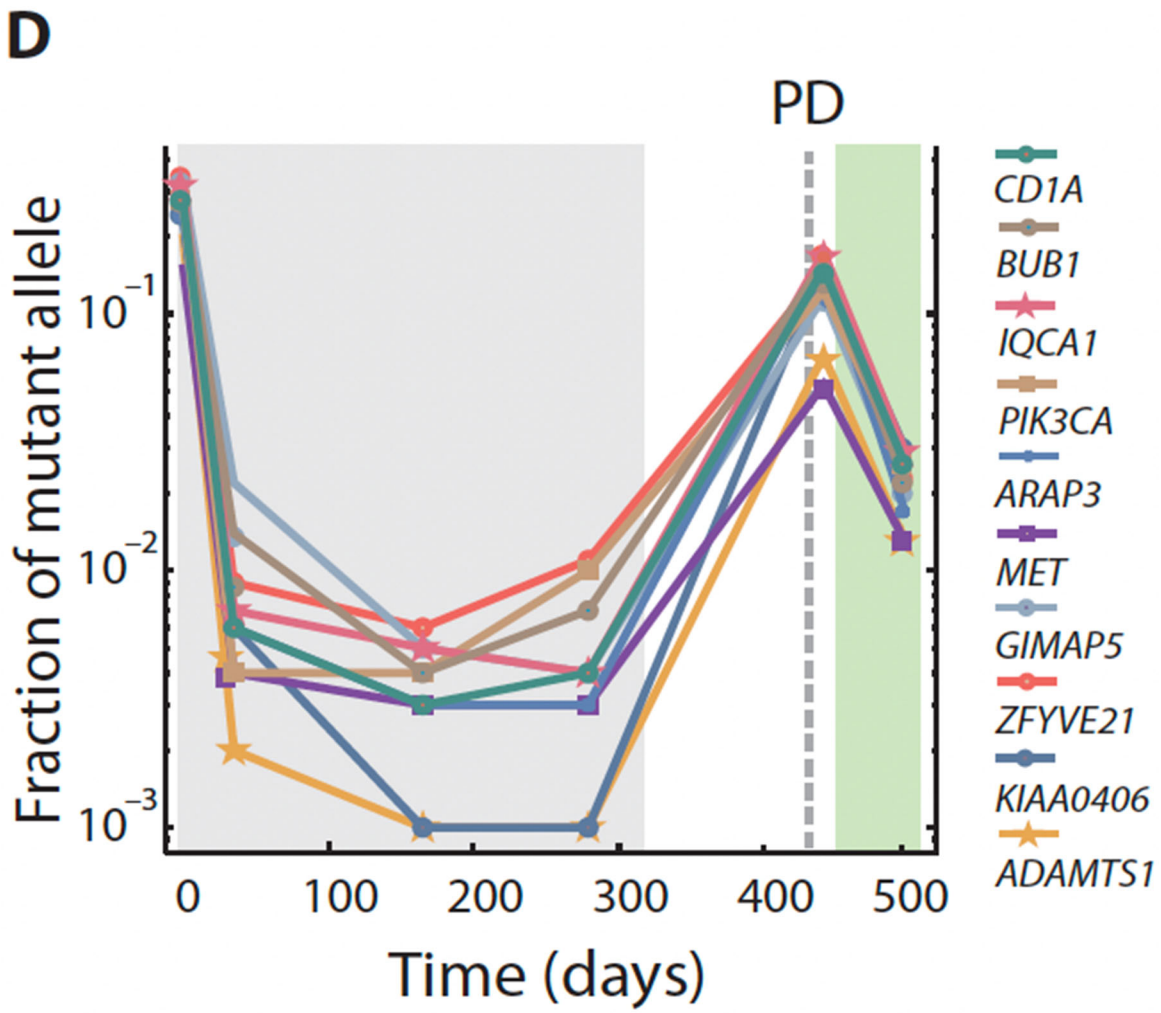
and pooled into a sequencing library (before a sequencing step). *See id.* at 2 (Fig. 1) (reproduced in part below).



c) "subjecting said sequencing library to sequencing to identify a presence or absence of at least a subset of said set of genetic variants in at least said additional biological sample from said subject, wherein a sequencing footprint of said sequencing of (c) is less than

a sequencing footprint of said sequence data of (a); and"

260. As noted above, the sequencing library created in the preceding step 1(b) is then subjected to sequencing to identify the presence or absence (specifically, the fraction of mutant allele) of the genetic variants in multiple plasma samples over time. This is shown in Figure 4D, which is reproduced below.



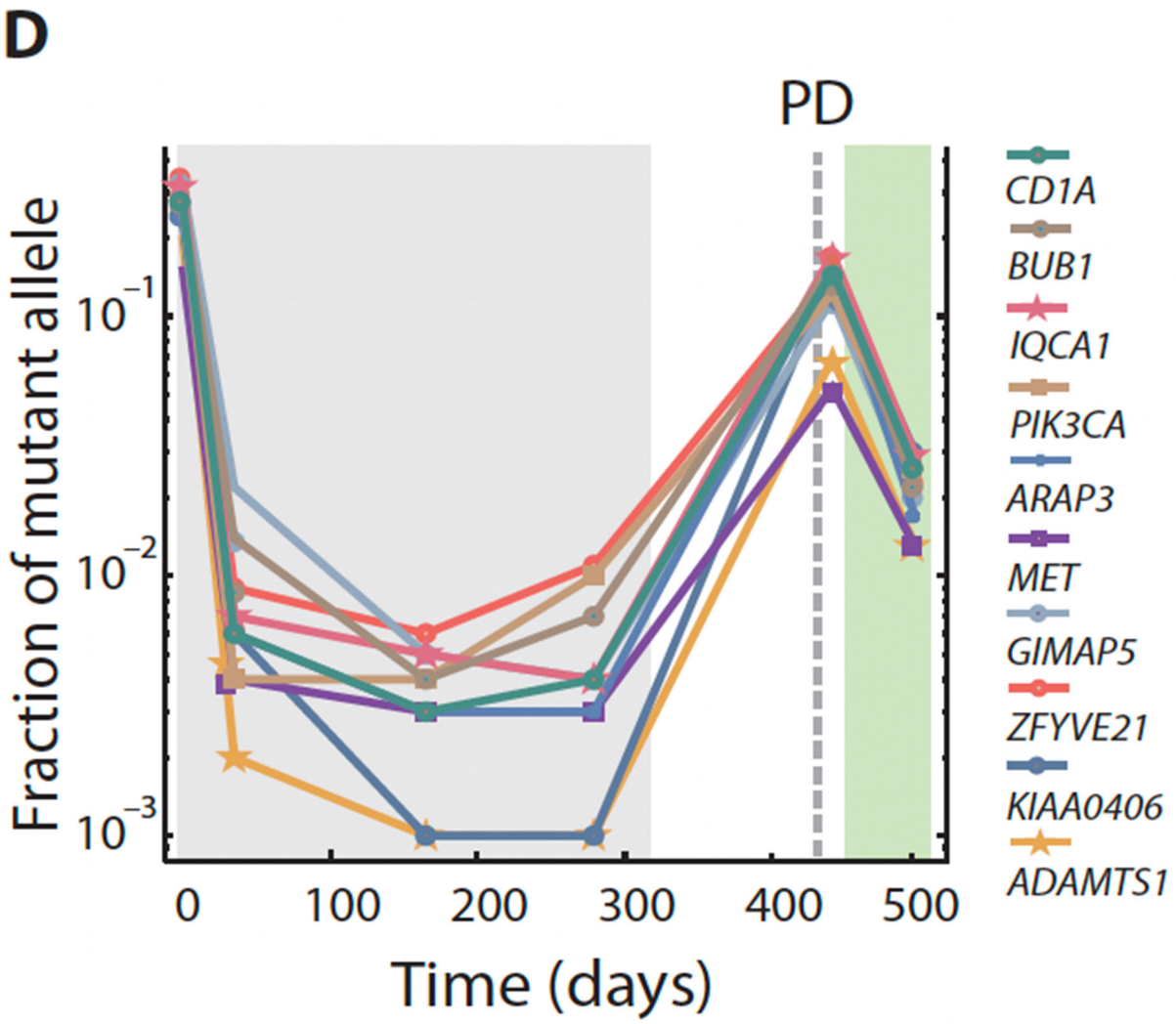
261. A POSA would have understood that the sequencing footprint of the targeted sequencing of 10 variant alleles (or even 11 with the addition of *TP53* probes according to the combination of Forshew and Dawson) will necessarily be less than the sequencing footprint of the whole-genome sequencing upon which the sequence data of step 1(a) is based.

262. Specifically, the sequencing footprint of the sequencing performed in Forshew's step "(c)" is less than a sequencing footprint of the "sequence data of (a)." The sequence data referenced in step (a) in Forshew is derived from whole-genome sequencing. This whole-genome sequencing data necessarily has a larger sequencing footprint than the targeted sequencing performed following amplification in this step (i.e., step 1(c)). In Forshew's methods, the sequencing of step (c) is performed using PCR amplicons that track the individual genes of interest identified in step (a), and thus are only small subsets of the whole genome. Accordingly, the sequencing footprint of the sequencing data of (c), which is generated by sequencing such amplicons, is necessarily less than the sequencing footprint of step (a)'s whole-genome sequence data. This is consistent with how Patent Owner reads this limitation onto the accused products. Ex. 1072 (Patent Owner's Infringement Contentions) at 4-5 (contending that limitation is met where variants "were tracked with probes of 120bp in length" while "the first sequencing step used to identify said variants (in the case of Foresight, whole genome sequencing)" has a larger footprint).

d) "generating a report identifying (i) a presence or absence of a health condition or disease of the subject in said additional biological sample based on said identified at least a subset of (c) and (ii) genetic variants which inform a therapy choice or a change in therapy for the subject."

263. Forshew further discloses generating a report that shows the presence or absence of a health condition or disease of the subject based on multiple plasma samples over time which inform a therapy choice or change in therapy for the subject.

264. Specifically, Figure 4D shows a report depicting the fraction of the various mutant alleles over time, including in response to therapy (grey box).



265. Notably, Forshew's Figure 4D "tracked dynamics of all mutations" analyzed with TAm-seq. *Id.* at 9. "The patient was treated with single-agent epirubicin (gray-shaded region). After 4 months off treatment, a CT scan showed progressive disease and the patient commenced further treatment with paclitaxel chemotherapy. The 10 mutations followed a common pattern of sharp decline in AF [allele fraction] upon onset of therapy and an increase in AF upon disease

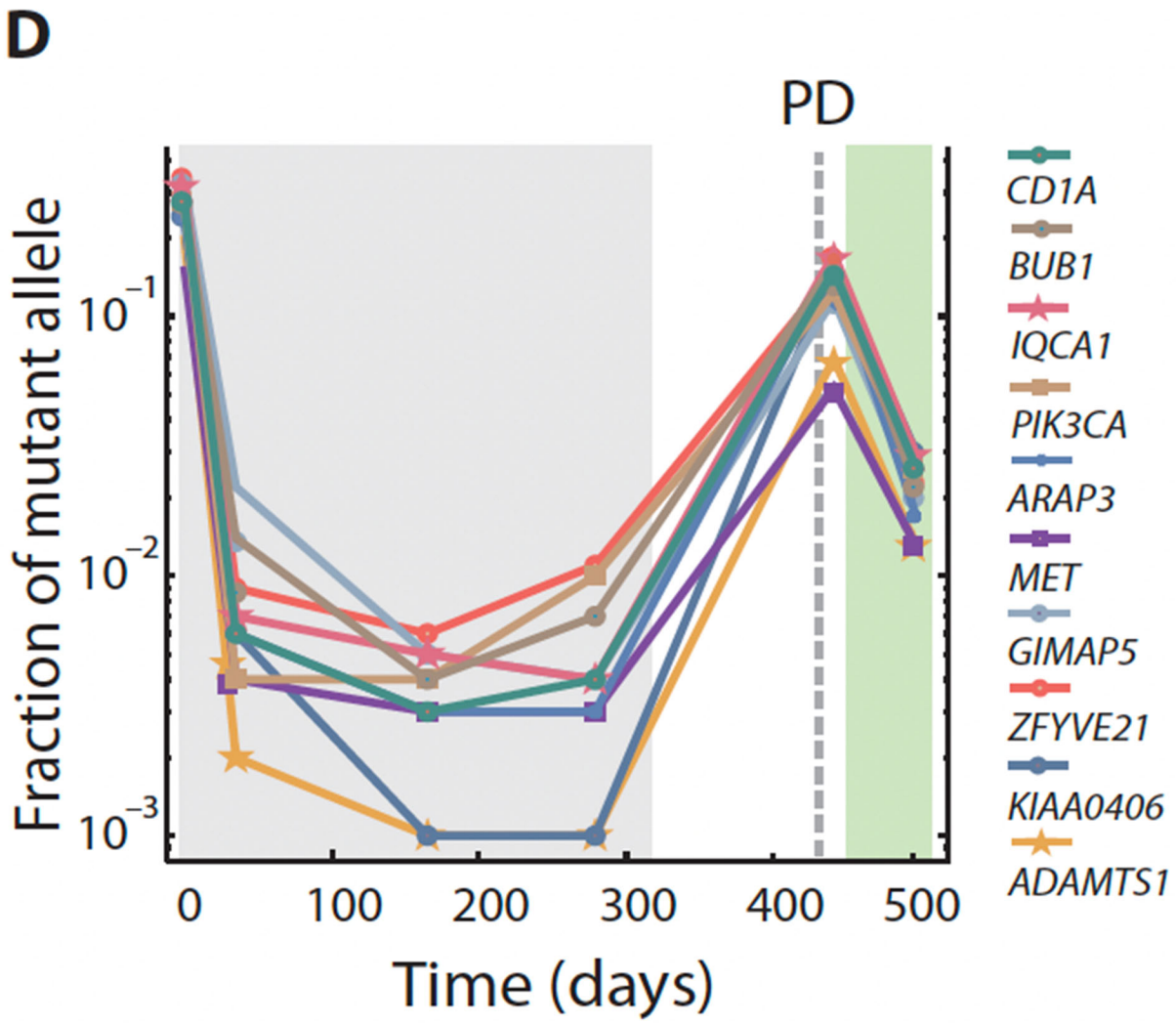
progression after termination of therapy." *Id.* Thus, Forshew's report informs therapy choice or a change in therapy based on the allele fraction of the identified alleles, including because of the observed relationship between therapy and disease status.

266. Accordingly, Forshew in view of Dawson renders obvious claim 1.

2. Claim 2

267. Claim 2 recites, "The method of claim 1, wherein said at least one biological sample comprises a plurality of biological samples."

268. As discussed above, Forshew discloses analyzing plasma samples taken at multiple time points over the course of nearly 500 days. *See* Ex. 1030 (Forshew) at 7 (Fig. 4(D)).



269. Accordingly, Forshew in view of Dawson renders claim 2 obvious.

3. Claim 3

270. Claim 3 recites, "The method of claim 2, wherein said plurality of biological samples comprises a tissue sample and a plasma sample."

271. As explained above, Forshew discloses that the plurality of samples comprise a tissue sample and a plasma sample. *See* Ex. 1030 (Forshew) at 9

("Whole-genome sequencing of ***tumor material*** was used to identify tumor mutations . . . . We used TAm-Seq . . . using DNA from 0.08 ml of plasma per amplification . . . .").

272. Accordingly, Forshew in view of Dawson renders claim 3 obvious.

4. Claim 4

273. Claim 4 recites, "The method of claim 2, wherein the plurality of biological samples comprise: (i) cell-free deoxyribonucleic acid molecules (cfDNA), (ii) cell-free ribonucleic acid molecules (cfRNA), (iii) DNA or RNA from circulating tumor cells, or (iv) RNA or DNA from a tumor metastasis."

274. Forshew discloses collecting plasma samples in "a patient with metastatic breast cancer." Ex. 1030 (Forshew) at 9. Such plasma samples necessarily include circulating cell-free DNA, RNA, and RNA or DNA from a tumor metastasis that was shed into the bloodstream. *See* Section V.E ("Related Applications of Next Generation Sequencing").

275. Accordingly, Forshew in view of Dawson renders claim 4 obvious.

5. Claim 5

276. Claim 5 recites, "The method of claim 2, further comprising generating said sequence data of (a) from the plurality of biological samples."

277. As explained above, Forshew describes performing whole-genome sequencing on a tissue biopsy from a breast cancer patient. A POSA would have

understood that to identify patient-specific mutations from whole-genome sequencing data, a standard practice is to take as an additional biological sample a control or normal tissue sample on which to perform whole-genome sequencing in order to provide a comparator for identifying tumor mutations. A POSA would recognize that this is almost certainly how Forshew's experiment was performed. By comparing whole-genome sequencing data for the tumor sample with normal tissue, sequence data identifying mutations of interest can be generated.

278. Ley illustrates this standard practice, as it discloses comparing the sequence data from the primary tumor sample to a sequenced skin sample, which constitutes a reference. *See* Ex. 1012 (Ley) at 68-69 ("Identification of somatic mutations in coding sequences"). Ley also describes comparing the sequence data to other references, such as previously described human genomes. *Id.*; *see also id.* at 67 ("Coverage depth of the tumour and skin genomes") ("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).").

279. Accordingly, Forshew in view of Dawson renders claim 5 obvious.

6. Claim 6

280. Claim 6 recites, "The method of claim 1, wherein said sequence data of (a) comprises sequence information for at least an exome, a panel of genes, or a whole genome."

281. As explained above, the sequence data of (a) in Forshew is from a whole genome.

282. Accordingly, Forshew in view of Dawson renders claim 6 obvious.

7. Claim 7

283. Claim 7 recites, "The method of claim 1, wherein identifying said presence or absence of a health condition or disease is based on a frequency of a genetic variant from said set of genetic variants."

284. Forshew explains that identifying the presence or absence of the patient's cancer is based on monitoring the allele fraction of the genetic variants over time: "The 10 mutations followed a common pattern of sharp decline in AF [allele fraction] upon onset of therapy and an increase in AF upon disease progression after termination of therapy (Fig. 4D)." Ex. 1012 (Forshew) at 9.

285. Accordingly, Forshew in view of Dawson renders claim 7 obvious.

8. Claim 8

286. Claim 8 recites, "The method of claim 1, wherein said additional biological sample of (c) is a blood sample."

287. As discussed above, Forshew discloses taking multiple additional biological samples over time. The samples taken in Forshew are plasma samples, which are blood samples.

288. Accordingly, Forshew in view of Dawson renders claim 8 obvious.

9. Claim 9

289. Claim 9 recites, "The method of claim 1, wherein said health condition or disease is cancer."

290. The health condition or disease addressed by Forshew is breast cancer.

291. Accordingly, Forshew in view of Dawson renders claim 9 obvious.

10. Claim 13

292. Claim 13 recites, "The method of claim 1, further comprising repeating (b) and (c) one or more times and generating one or more additional reports."

293. As explained above, Forshew discloses repeating steps (b) and (c) multiple times on multiple plasma samples in order to track allele fraction over time. In Figure 4D, Forshew provides a report of all the data monitored over the course of its study. A POSA would have found it obvious to generate such a report at each time point, adding the subsequently obtained information at each additional time point following data collection.

294. Accordingly, Forshew in view of Dawson renders claim 13 obvious.

11. Claim 14

295. Claim 14 recites, "The method of claim 1, wherein said at least one additional biological sample in (c) was obtained after said at least one biological sample in (a)."

296. Forshew's Figure 4D shows that multiple additional biological samples are obtained at multiple time points following the initial sample over the course of nearly 500 days.

297. Accordingly, Forshew in view of Dawson renders claim 14 obvious.

12. Claim 15

298. Claim 15 recites, "The method of claim 1, wherein said additional biological sample is a blood sample comprising: (i) cfDNA, (ii) cfRNA, (iii) DNA or RNA from circulating tumor cells, or (iv) RNA or DNA from a tumor metastasis."

299. Forshew discloses collecting multiple plasma samples in "a patient with metastatic breast cancer." Ex. 1030 (Forshew) at 9. Such plasma samples necessarily include circulating cell-free DNA, RNA, and RNA or DNA from a tumor metastasis that was shed into the bloodstream. *See* Section V.E ("Related Applications of Next Generation Sequencing").

300. Accordingly, Forshew in view of Dawson renders claim 15 obvious.

13. Claim 16

301. Claim 16 recites, "The method of claim 1, wherein said additional biological sample and one or more subsequent biological samples are each blood samples comprising: (i) cfDNA, (ii) cfRNA, (iii) DNA or RNA from circulating tumor cells, or (iv) RNA or DNA from a tumor metastasis."

302. Forshew discloses collecting multiple plasma samples in "a patient with metastatic breast cancer." Ex. 1030 (Forshew) at 9. Such plasma samples necessarily include circulating cell-free DNA, RNA, and RNA or DNA from a tumor metastasis that was shed into the bloodstream. *See* Section V.E.

303. Accordingly, Forshew in view of Dawson renders claim 16 obvious.

14. Claim 17

304. Claim 17 recites, "The method of claim 1, wherein said at least one fixed portion comprises one or more known tumor variants."

305. The fixed portion of Forshew in view of Dawson comprises probes for *TP53.* As discussed above, *TP53* is known to be a mutate variant tumor suppressor gene present in breast cancers.

306. Accordingly, Forshew in view of Dawson renders claim 17 obvious.

15. Claim 18

307. Claim 18 recites, "The method of claim 1, wherein said at least one fixed portion is selected from a group consisting of one or more of: (i) cancer driver

genes, (ii) genes involved in the pharmacogenomics of cancer drugs, (iii) genes involved in Mendelian immunological diseases, (iv) genes related to inherited forms of cancer, (v) genes associated with tumor escape from a targeted or immune cancer therapy, (vi) HLA typing, (vii) variants common in the population and used by B-allele methods to detect structural variation, and any combination thereof."

308. As discussed above, *TP53* genes included in the fixed panel rendered obvious by Forshew in view of Dawson constitute cancer driver genes (through presence or absence of tumor suppression) and are genes related to inherited forms of cancer.

309. Accordingly, Forshew in view of Dawson renders claim 18 obvious.

16. Claim 19

310. Claim 19 recites, "The method of claim 1, wherein said sequence data of (a) is derived from sequences of nucleic acids from said at least one biological sample, wherein said at least one biological sample comprises a tumor biopsy."

311. As discussed above, Forshew's initial whole-genome sequencing data is derived from sequences of nucleic acids taken from a tumor sample, i.e., a tumor biopsy.

312. Accordingly, Forshew in view of Dawson renders claim 19 obvious.

17. Claim 21

313. Claim 21 recites, "The method of claim 1, wherein said set of one or more genetic variants comprise one or more members selected from the group consisting of: (i) a single nucleotide polymorphism, (ii) an insertion or deletion, (iii) a copy number variation, and (iv) structural variation."

314. Given the types of mutations monitored in Forshew (which include SNPs) a POSA would have found it obvious to include probes for variants constituting SNPs, insertions or deletions, or structure variations, given that those represent the most common types of cancer mutations. Indeed, multiple other reference monitor such types of variations using similar assays. *See* Ex. 1012 (Ley); Ex. 1002 (Leary); Ex. 1015 (McBride).

315. Accordingly, Forshew in view of Dawson renders claim 21 obvious.

18. Claim 22

316. Claim 22 recites, "The method of claim 1, wherein said personalized probe set comprises hybrid capture probes."

317. Forshew discloses the use of PCR primers as probes for the amplification and/or enrichment of target DNA over normal DNA in the sample to prepare its sequencing library. However, by the 2016 priority date of the '783 patent, a POSA would have also been familiar with the use of hybrid capture techniques for similar purposes. As explained in Section V.C.3, above, PCR and hybrid capture

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"). Accordingly, a POSA would have found it obvious to use create hybrid capture probes targeting the same sequences as Forshew in view of Dawson because doing so would have involved only the predictable usage of techniques known in the art to be interchangeably useful for the same purpose. Indeed, 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 2011) at 379. Thus, a POSA would have readily been able to obtain "out of the box" hybrid capture reagents for a "custom enrichment region" including a panel of probes directed to the genetic variations one would monitor in Forshew's methods and the variations thereof discussed in this section.

318. The '783 patent's specification confirms that those skilled in the art considered these techniques to be interchangeable as of the priority date. In defining the terms "probe," "nucleic acid probe," "capture probe," and "bait," the '783 patent explains:

> The terms "probe," "nucleic acid probe," "capture probe," "bait," as used herein, generally refer to a nucleic acid molecule comprising a single-stranded portion capable of hybridizing to a complementary nucleic acid sequence. A probe can be used for detection or enrichment of nucleic acid molecules. A probe can be target-specific such that a region of interest may be pulled-down, isolated, enriched, amplified, or labeled. A probe can be used for targeted sequencing. A probe may hybridize to a targeted sequence when attached to a solid substrate or when in-solution, e.g. as for hybrid capture. A probe may be included in a set, or plurality, of probes. A probe set can comprise probes that overlap within a specific genomic region such that they are tiled or staggered. A probe set can include probes to a genomic region or a panel comprising multiple genomic regions. Probes can be amplification based or capture hybridization-based. Non-limiting examples of probes include molecular inversion probes, amplification probes, biotinylated affinity probes, or any probe comprising a detectable label.

Ex. 1001 at 29:4-22.

319. Accordingly, Forshew in view of Dawson and a POSA's knowledge renders claim 22 obvious.

19. Claim 23

320. Claim 23 recites, "The method of claim 1, wherein said set of genetic variants identified in sequence data are determined with respect to a reference."

321. As explained above, Forshew describes performing whole-genome sequencing on a tissue biopsy from a breast cancer patient. Although not expressly disclosed in Forshew, a POSA would have found it obvious to take as an additional biological sample a control or normal tissue sample on which to perform whole-genome sequencing in order to provide a comparator (or reference) for identifying tumor mutations. By comparing whole-genome sequencing data for the tumor sample with normal tissue, sequence data identifying mutations of interest can be generated.

322. Ley illustrates this standard practice, as it discloses comparing the sequence data from the primary tumor sample to a sequenced skin sample, which constitutes a reference. *See* Ex. 1012 (Ley) at 68-69 ("Identification of somatic mutations in coding sequences"). Ley also describes comparing the sequence data to other references, such as previously described human genomes. *Id.*; *see also id.* at 67 ("Coverage depth of the tumour and skin genomes") ("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).").

323. Accordingly, Forshew in view of Dawson renders claim 23 obvious.

20. Claim 24

324. Claim 24 recites, "The method of claim 1, further comprising, prior to (a), sequencing said least one biological sample to generate said sequencing data."

325. As explained above, Forshew discloses performing whole-genome sequencing on the primary tumor sample prior to step 1(a) in order to generate said sequencing data.

326. Accordingly, Forshew in view of Dawson renders claim 24 obvious.

21. Claim 25

327. Claim 25 recites, "The method of claim 1, further comprising generating said personalized probe set for said subject based on said set of genetic variants identified in said sequence data for said at least one biological sample."

328. As discussed above, the personalized probe set disclosed in Forshew is based on the sequencing data identified in the primary tumor using whole-genome sequencing. Specifically, Forshew discloses including primers for the mutations identified based on the whole-genome sequencing of the breast cancer patient's tumor.

329. Accordingly, Forshew in view of Dawson renders claim 25 obvious.

22. Claim 26

330. Claim 26 recites, "The method of claim 1, wherein (b) further comprises using said personalized probe set to selectively enrich or amplify

sequences comprising said set of genetic variants over other sequences in a biological sample from another subject to generate said sequencing library."

331. As explained above, applications of similar sequencing technologies were known to those skilled in the art to be applicable to the monitoring of cell-free DNA in various contexts, including organ transplant rejection and prenatal diagnostics. *See* Section V.E ("Related Applications of Next Generation Sequencing"). In those scenarios, when there are foreign cells present in the body – whether from a fetus or from a donor organ – those cells release into the blood stream cell free DNA with its own signature. In the context of a cancer patient that was also pregnant or the recipient of an organ transplant, a POSA would have found it obvious to selectively enrich or amplify sequences comprising the genetic variants of the patient's cancer over other sequences that may be present in the sample from another subject – such as circulating fetal DNA or organ donor-derived DNA – in order to generate the sequencing library. A POSA would have been motivated to select sequences unique to the tumor, as opposed to the foreign DNA from a fetus or donor organ, in order to precisely monitor the cancer over other medical issues of concern.

332. Accordingly, Forshew in view of Dawson and the knowledge of a POSA renders claim 26 obvious.

23. Claim 27

333. Claim 27 recites, "The method of claim 1, wherein (c) further comprises identifying levels of gene expression, sequencing read counts, or read depth."

334. Forshew renders obvious identifying sequencing read counts and read depth. For example, Forshew explains that to detect mutations at lower frequencies, a higher number of sequencing read counts, and thus a greater read depth, should be utilized. Ex. 1030 at 3 ("The minimal detectable frequency would also be limited if a minimal number of reads is applied for confident mutation detection; for example, a minimum of 10 reads implies that sequencing depth of 5000 would be required to detect mutations at AF as low as 0.2%. For alleles present at ~10 or fewer copies in the starting template, reproducibility would also be limited by sampling noise, because these alleles may be over- or underrepresented in any particular reaction.").

335. Accordingly, Forshew in view of Dawson renders claim 27 obvious.

24. Claim 28

336. Claim 28 recites, "The method of claim 1, wherein said at least one additional biological sample comprises a plurality of additional samples obtained at different time points."

337. As explained above, the additional biological samples in Forshew comprise multiple plasma samples taken at multiple additional time points over the

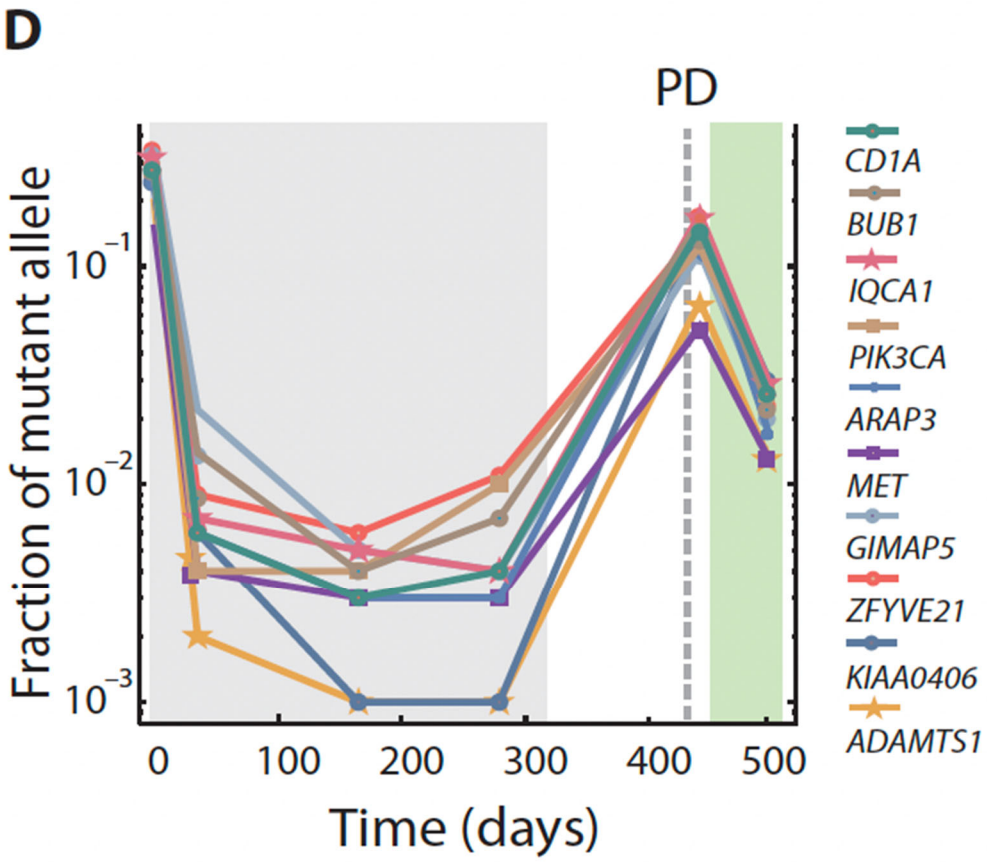
course of nearly 500 days following the initial sample collection and analysis. *See* Ex. 1030 (Forshew) at 7 (Fig. 4D).

338. Accordingly, Forshew in view of Dawson renders claim 28 obvious.

25. Claim 29

339. Claim 29 recites, "The method of claim 1, wherein the report of (d) further comprises reporting biomedical information of said subject that is predictive, prognostic, or diagnostic of one or more biomedical features selected from the group consisting of a disease state, efficacy of a drug therapy, prediction of optimal drug dosage, a recommendation of one or more therapies, and a recommendation of a course of treatment of a disease."

340. The report of (d) disclosed in Forshew's Figure 4(d) provides information regarding the fraction of mutant alleles over time, including during the time period when the patient is receiving therapy (gray box).



341. Thus, the report provided in Forshew is at least prognostic and/or diagnostic regarding the efficacy of a drug therapy. Specifically, Forshew shows where the therapy is working (reduction in allele fraction) and also provides information regarding the time point at which the cancer appears to begin resisting the therapy, where the allele fraction begins to increase near day 300.

342. Accordingly, Forshew in view of Dawson renders claim 29 obvious.

IX. Conclusions

343. For the reasons set forth above, the Challenged Claims are anticipated and/or rendered obvious by the prior art.

IPR2023-00545
U.S. Patent No. 11,299,783

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: By:_____ Digitally signed by John Quackenbush Date: 2023.02.10 13:08:45 -05'00'

John Quackenbush, Ph.D.