

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ILLUMINA INC.,
Petitioner,

v.

MOLECULAR LOOP BIOSCIENCES, INC.,
Patent Owner.

Case No. IPR2024-00965
Patent No. 11,768,200

DECLARATION OF PAUL T. SPELLMAN, PH.D.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
A. Qualifications and Experience	1
B. Materials Considered.....	6
II. LEVEL OF ORDINARY SKILL IN THE ART.....	6
III. SUMMARY OF OPINIONS.....	7
IV. LEGAL STANDARDS	13
V. BACKGROUND AND STATE OF THE ART.....	16
A. First Generation Sequencing	16
B. Next Generation Sequencing Techniques Available as of December 2010.....	19
1. Overview of Sequencing Workflow	20
2. Basic Principles Behind Illumina and Roche 454 Sequencing	22
3. Immobilization of Nucleic Acids of Interest	26
4. Amplification of Nucleic Acids of Interest.....	27
5. Generation of Sequence Reads and Sequences of Nucleic Acids of Interest.....	30
6. Use of Sequence Identifiers in Multiplex Sequencing	32

7.	Resources available to a POSA before December 2010.....	40
VI.	SUMMARY OF THE PRIOR ART.....	41
A.	<i>Parameswaran</i> (EX1004)	41
B.	<i>Gloor</i> (EX1005)	47
C.	Binladen (EX1029).....	51
D.	<i>Bentley</i> (EX1006).....	54
VII.	THE '200 PATENT	57
A.	Overview	57
B.	The Claims of the '200 Patent.....	57
C.	Disclosures of the '200 Patent.....	59
D.	Prosecution History	65
1.	Prosecution History of the '281 Patent.....	67
2.	Prosecution History of the '200 Patent.....	68
E.	Claim Construction.....	69
VIII.	THE CHALLENGED CLAIMS ARE UNPATENTABLE OVER THE PRIOR ART	69
A.	Ground 1: Claims 1-6 Are Anticipated by <i>Gloor</i>	69
1.	Claim 1 is Anticipated by <i>Gloor</i>	69
2.	Dependent Claims 2-6 Are Anticipated by <i>Gloor</i>	78
B.	Ground 2: Claims 1-6 Are Anticipated by <i>Parameswaran</i>	81

1.	Claim 1 Is Anticipated by <i>Parameswaran</i>	81
2.	Dependent Claims 2-6 Are Anticipated by <i>Parameswaran</i>	91
C.	Ground 3: Claims 1-6 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Gloor</i>	94
1.	Claim 1 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Gloor</i>	95
2.	A POSA Would Have Been Motivated to Combine <i>Parameswaran</i> and <i>Gloor</i>	98
3.	A POSA Would Have Had a Reasonable Expectation of Success in Combining <i>Parameswaran</i> and <i>Gloor</i>	103
4.	Dependent Claims 2-6 Would Have Been Obvious Over the Combination of <i>Parameswaran</i> and <i>Gloor</i>	107
D.	Ground 4: Claims 1-6 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Binladen</i>	109
1.	Claim 1 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Binladen</i>	110
2.	A POSA Would Have Been Motivated to Combine <i>Parameswaran</i> and <i>Binladen</i>	117

3.	A POSA Would Have Had a Reasonable Expectation of Success in Combining <i>Parameswaran</i> and <i>Binladen</i>	119
4.	Dependent Claims 2-6 Would Have Been Obvious Over the Combination of <i>Parameswaran</i> and <i>Binladen</i>	120
E.	Ground 5: Claims 6 Would Have Been Obvious Over <i>Parameswaran</i> , <i>Binladen</i> , and <i>Bentley</i>	122
F.	No Unexpected Results or Other Evidence of Nonobviousness.....	125
IX.	CONCLUSION.....	126

I. INTRODUCTION

1. I, Dr. Paul T. Spellman, have been retained as an independent expert in the field of genetics and DNA sequencing. I submit this declaration on behalf of Petitioner Illumina, Inc., in the above-captioned *inter partes* review (“IPR”).

2. I am being compensated for my time in connection with this IPR at my standard hourly consulting rate of \$500/hour. I do not have any personal or financial stake or interest in the outcome of this proceeding and my compensation is in no way contingent on the nature of my analysis or the outcome of this IPR or any other proceeding.

3. I am over 21 years of age and, if I am called upon to do so, I would be competent to testify as to the matters set forth herein.

A. Qualifications and Experience

4. I believe that I am qualified to serve as a technical expert in this matter based upon my qualifications, discussed in detail below. A copy of my curriculum vitae is attached as **Appendix A** to this declaration.

5. I am a Professor of Medicine in the Division of Hematology and Oncology as well as in the Department of Human Genetics at the University of California Los Angeles (“UCLA”) David Geffen School of Medicine. My research focuses on bioinformatics and genome sequencing, including developing improved sequencing techniques and applying genomic and computational technologies to improve human health. My research encompasses all phases of genomic research and sequencing, from technology and method development to application of technologies to answer critical questions in cancer biology, to population studies to understand the impact of genetic variation of disease, and to implementation trials that directly impact health. I have more than 25 years of experience in nucleic acid sequencing and have first-hand experience in designing multiplex sequencing assays.

6. I graduated from the Massachusetts Institute of Technology in 1995 with a Bachelor of Science in Biology. I then earned my Ph.D. in Genetics from the Stanford University School of Medicine in 2000.

7. Following my Ph.D., I conducted post-doctoral research from 2000 to 2003 at the University of California Berkley Department of Molecular and Cellular Biology, studying gene regulation and genomics. From 2003 to 2011, I was a Scientist and Staff Scientist at the Lawrence Berkeley Lab, Life Science Division. In 2011, I also served as the Special Assistant to the Deputy Director of the National Cancer Institute.

8. In 2011, I joined the Oregon Health & Science University (“OHSU”) School of Medicine first serving as an Associate Professor, then Professor with tenure, and then the Penny and Phil Knight Endowed Professor in Cancer Research Innovation. There, I studied the use of population genetics to help determine who is at risk for cancer, how to computationally analyze genomic data to identify early changes in cancers, and how to accurately screen different populations for the disease. My research also included using genetic and genomic approaches to understand the processes by which cancer develops, monitor disease, and identify therapeutic strategies.

9. I spent 12 years at the OHSU School of Medicine before joining the faculty of UCLA in 2023. During that time at OHSU, I held multiple leadership roles for programs relating to genomic sequencing, including serving as the Program Leader of the Quantitative Oncology Program, the Co-Director of the Cancer Early Detection Advanced Research (“CEDAR”) Center in the OHSU Knight Cancer Institute, and the Interim Director for the Program in Computational Biology.

10. To date, I have authored about 200 publications, including 175 articles, 3 book chapters, 5 letters to the editor, 12 literary reviews, and 4 abstracts, many of which relate to genomic research or apply genomic sequencing.

11. Since 2002, I have served on 25 grant review committees, 6 conference planning committees, and over 30 project advisory, institutional, and editorial committees. I am also active in related professional societies, including having served as Treasurer and Board member for the Microarray Gene Expression Data Society and Data Coordination and as a member of the Management Working Group and Scientific Planning Committee for the International Cancer Genomics Consortium.

12. I have served as a regular reviewer for journals relating to genome sequencing and biotechnology, including Bioinformatics, Nucleic Acids Research, Nature Communications, PLoS ONE, Cell Reports, Genome Research, and Genes, among others.

13. Since 2004, I have been awarded dozens of grants, fellowships, and other sources of funding for my work, including for studying topics such as developing cost effective sequence-based technologies, studying genes relating to breast and ovarian cancer, and analyzing genomic data. Currently, I have research funding for a systematic analysis of genetic and gene regulation information in clinical cohorts as part of the Genome Data Analysis Network and a clinical trial implementing genetic health screening for hereditary breast and ovarian cancer and Lynch syndromes.

14. Over the years, my work has been highlighted by news media outlets, including by the Salem Statesman Journal. I have given over 80 guest lectures on my work, and I have taught over 50 courses, workshops, and seminars on topics including genetics, genome sequencing, biotechnology, and sequencing technologies. I have also served as an advisor to over 20 post-doctoral fellows and pre-doctoral graduate students.

15. As high-throughput sequencing has come to dominate the technological basis of my career, I have been well aware of, and my lab routinely uses, unique and non-unique dual-index sequencing methods.

B. Materials Considered

16. My analysis in this declaration is based on my education, personal knowledge, and professional and academic experience in the areas of genetics and DNA sequencing, including multiplex sequencing and dual-indexing methods. My analysis is particularly focused on the state of the art in December of 2010.

17. In forming my opinions, I have considered the '200 patent and other exhibits cited in my declaration, including those listed in **Appendix B**. I reserve the right to rely on documents cited in the appendices to this declaration, and to supplement my opinions in view of new materials and information that becomes available to me during this proceeding.

II. LEVEL OF ORDINARY SKILL IN THE ART

18. I understand that an assessment of a patent's claims should be undertaken from the perspective of a person of ordinary skill in the art ("POSA") as of the "effective filing date" of the patent claims. In performing my analysis, I have been asked to assume that the "effective filing date" of the '200 patent is December 23, 2010.

19. I understand that the level of ordinary skill in the art is determined by looking at; (1) the type of problems encountered in the art; (2) the prior-art solutions to those problems; (3) the rapidity with which innovations are made; (4) the sophistication of the technology; and (5) the educational level of active workers in the field.

20. In my opinion, a POSA would have had (i) a Ph.D. in molecular biology, genetics, bioinformatics, or a related field and at least two years of experience in high-throughput sequencing technologies or (ii) a Master's degree in one of the same fields with at least four years of experience in high-throughput sequencing technologies. As explained above in Section I.A, I had at least these qualifications by December 2010 and am qualified based on my education and experience to provide an opinion as to what a POSA would have known and concluded as of December 2010. *Supra* § I.A.

III. SUMMARY OF OPINIONS

21. I understand that this declaration accompanies a Petition for IPR involving U.S. Patent No. 11,768,200 (“the ’200 patent”) (EX1001). I understand that the earliest possible “effective filing date” for the ’200 patent is December 23, 2010. In performing my analysis, I have been asked to assume that the “effective filing date” of the ’200 patent is December 23, 2010.

22. The claims of the '200 patent are directed to a method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction. As discussed below, the sequencing workflow typically involved several steps, including extracting template nucleic acids from samples, preparing those template nucleic acids for sequencing, and sequencing the products. It was often advantageous to use polymerase chain reaction (PCR) or other amplification methods to replicate the template during the sequencing workflow to increase the number of molecules available for sequencing. In the case of multiplex sequencing where multiple samples were sequenced simultaneously, sequencing products were combined ("pooled") at various steps during the sequencing workflow to gain efficiencies and reduce costs, including for PCR or other forms of DNA amplification.

23. Numerous different types of errors, however, could occur during this workflow. The original template could get damaged during extraction or amplification. Primers used for DNA amplification could be erroneously synthesized. Further, during DNA amplification, chimeric molecules consisting of two template nucleic acids could be formed. When these errors occurred in multiplex sequencing, the accuracy of sequencing could be significantly affected, resulting in erroneous sequences that did not exist in the original samples or misassignments of sequences to incorrect samples.

24. The claims of the '200 patent purport to address these known problems by using a dual-indexing method. The claims recite that the sequence of a nucleic acid analyte of interest is validated by detecting the presence of two or more identifier sequences uniquely associated with the nucleic acid analyte of interest, sequencing the nucleic acid analyte of interest flanked with the two identifier sequences, and validating the sequence by analyzing both identifier sequences and excluding those sequences containing only one or an incorrect pair of identifier sequences.

25. Based on my knowledge and experience and my review of the information described in this declaration, it is my opinion that claims 1-6 of the '200 patent are anticipated by two references, each of which disclosed every step of the claimed dual-indexing methods.

26. First, *Gloor* et al., Cornell University Library arXiv:1007.5075v1 (2010) ("*Gloor*"; EX1005), anticipates claims 1-6 of the '200 patent. *Gloor* disclosed a multiplex sequencing method where nucleic acid analytes from multiple samples were tagged with a pair of identifier sequences. While *Gloor* repeated individual identifier sequences across multiple samples, the pairs of identifiers did not repeat, i.e., each pair of identifiers was uniquely associated with nucleic analytes of interest from a single sample. Using Illumina sequencing, *Gloor* sequenced nucleic acid analytes from multiple samples simultaneously. *Gloor* also validated

the sequences of the analytes by analyzing the two identifiers present in each sequence, and excluding sequences that contained only one identifier or pairs of incorrect identifiers. Because *Gloor* discloses each limitation of the challenged claims, *Gloor* anticipates those claims.

27. *Parameswaran* et al., Nucleic Acids Research (2007) 35:e130 (“*Parameswaran*”; EX1004) also anticipates claims 1-6 of the ’200 patent. Like *Gloor*, *Parameswaran* disclosed multiplex sequencing methods where nucleic acid analytes of interest from each sample were tagged with a unique pair of identifier sequences. Using Roche 454 sequencing, *Parameswaran* sequenced nucleic acid analytes from multiple samples simultaneously. *Parameswaran* validated the sequences of the analytes by analyzing the two identifiers and separated sequences with only one identifier from those with two correct identifiers. Among those separated sequences, *Parameswaran* identified sequences containing incorrect pairs of identifiers and characterized them as “false-discover[ies]” constituting “misassignment[s]” of sequences. Because *Parameswaran* discloses each limitation of the challenged claims, *Parameswaran* anticipates those claims.

28. In addition, it is my opinion that claims 1-6 of the ’200 patent would have been obvious to a person of ordinary skill in the art (“POSA”) as of December 23, 2010, over the combination of *Parameswaran* and *Gloor*.

29. Each and every element of the claims of the '200 patent is found in the combination of *Parameswaran* and *Gloor*. As discussed below, a POSA would have been motivated to combine *Parameswaran* and *Gloor* with a reasonable expectation of success. Indeed, to maximize sequencing accuracy, cost, and efficiency, a POSA would have been motivated to use *Parameswaran*'s dual-indexing method, which identified different types of erroneous sequences produced during multiplex sequencing, with the Illumina sequencing used in *Gloor*. A POSA would have been motivated to exclude those erroneous sequences from further analysis to achieve "more accurate" Illumina sequencing. A POSA would have had a reasonable expectation of success in doing so because both *Gloor* and *Parameswaran* empirically demonstrated that erroneous sequences generated during a multiplex sequencing reaction can be readily identified using dual-indexing methods and excluded to improve accuracy.

30. It is also my opinion that claims 1-6 of the '200 patent would have been obvious to a POSA in view of *Parameswaran* and *Binladen* et al., PLoS ONE (2007) 2(2):e197 ("*Binladen*"; EX1029), another reference that used the Roche 454 platform to sequence multiple samples simultaneously, and identified and discarded erroneous sequences using a dual-indexing method. To improve the accuracy of multiplex sequencing, a POSA would have been motivated to use the dual-indexing methods in *Parameswaran* and *Binladen* to identify and exclude erroneous

sequences. A POSA would have had a reasonable expectation of success in combining *Parameswaran* and *Binladen* because both references demonstrated that the presence or absence of dual indexes in sequences can be used to identify and exclude erroneous sequences, thereby lowering sequence misassignment rates.

31. Claim 6 of the '200 patent recites that the nucleic acid analytes of interest are on “on the surface of a flow cell” during the multiplex sequencing reaction. To the extent this requires direct attachment of the analytes to the surface of the flow cell, claim would have been also obvious to a POSA over the combination of *Parameswaran*, *Binladen*, and *Bentley* et al., Nature (2008) 456:53-59 (“*Bentley*”; EX1006). *Bentley* teaches Illumina sequencing on the surface of a flow cell. EX1006 (*Bentley*), 2 (Figure 1). While *Parameswaran* and *Binladen* used Roche 454 sequencing, a POSA would have been motivated to use Illumina sequencing, as taught in *Bentley*, which was known to be faster, more efficient, and less expensive than Roche 454 sequencing. A POSA would have a reasonable expectation of success in making that switch in view of *Bentley*’s detailed guidance for performing Illumina sequencing.

32. Thus, claims 1-6 of the '200 patent would have been obvious to a POSA as of December 23, 2010. I am currently not aware of any objective evidence of non-obviousness for the claims of the '200 patent. I reserve the right to consider and comment on any such evidence after I sign this declaration.

IV. LEGAL STANDARDS

33. The opinions I express in this declaration involve the application of my technical knowledge and experience in evaluating certain prior art with respect to the '200 patent. In preparing this declaration, certain patent law concepts have been explained to me by counsel, including the legal standard for interpreting claims as well as those for assessing obviousness of a patent claim.

34. I have been informed that, in IPR proceedings such as this one, the party challenging the patent bears the burden of proving unpatentability by a preponderance of the evidence. I understand that a preponderance of the evidence means “more likely than not.”

35. I understand that patentability must be analyzed from the perspective of a POSA in the same field as the challenged patent as of the “effective filing date” of the claims. I understand that a POSA is a hypothetical individual presumed to know the relevant art as of the effective filing date of the claims.

36. I understand that the Patent Trial and Appeal Board interprets claims based on their ordinary meaning as understood by a POSA at the time of the effective filing date in view of the claim language, patent specification, and prosecution history. I also understand that the specification may reveal a special definition given to a claim term by an inventor that differs from its ordinary meaning as understood by a POSA. In that case, I understand that the inventor’s definition governs.

37. I understand that a patent claim is invalid as *anticipated* if the claimed subject matter was described in a printed publication before the effective filing date. I have been informed that a patent claim is anticipated if a single prior art reference discloses all limitations of the claimed subject matter. I have been informed that the anticipatory prior art reference need not use the very same terms used in the patent claim. I also understand that a reference disclosing a method of using one composition within the scope of the claimed subject matter anticipates the claim.

38. I understand that a patent claim is invalid if it would have been *obvious* to a person of ordinary skill in the art as of the effective filing date. I understand that assessing obviousness entails considering: (1) the scope and content of the prior art, (2) the differences between the prior art and the claimed invention, (3) the level of ordinary skill in the art, and (4) any secondary considerations of non-obviousness.

39. I understand that a claim may be obvious based on a combination of multiple prior-art references, as well as based on the knowledge and skill of a POSA as of the effective filing date. I also understand that to combine prior-art references, there must have been a motivation that would have prompted a POSA to do so. In other words, obviousness requires a *motivation to combine* the features of the prior art. I further understand that a POSA must have reasonably expected that the combination would work. That is, obviousness requires a *reasonable expectation of success* in combining the prior art to achieve the claimed subject matter. I also

understand that, while obviousness requires a reasonable expectation of success, it does not require absolute predictability of success in achieving the claimed subject matter.

40. I understand that a motivation to combine prior-art references may arise from a variety of sources, including, among other things, scientific literature, a need or unsolved problem in the field, or market demand.

41. I also understand that a claim may be obvious if a POSA would have found it obvious to try combining a finite number of predictable solutions known in the art. For example, if one technique had been used to improve a method of using a device, and a POSA would recognize that the same technique was one of a limited number of solutions and would improve similar devices in the same way, a POSA may have found it obvious to try that same technique on similar devices.

42. I understand that, when present, evidence of “secondary considerations” must be considered along with the other factual evidence relevant to obviousness. Such secondary considerations could include unexpected results, commercial success of products or processes using the invention, long-felt but unmet need for the invention, failure of others to make the invention, industry acceptance of the invention, or copying of the invention by others. I understand that, in order for such evidence to support nonobviousness, it must have a “nexus” to the features of the invention as claimed.

43. I have been informed that a dependent claim is a patent claim that refers back to another patent claim. I have been informed that a dependent claim includes all of the limitations of the claim to which it refers.

V. BACKGROUND AND STATE OF THE ART

A. First Generation Sequencing

44. Deoxyribonucleic acid (“DNA”) carries genetic information. DNA consists of two “reverse complementary” strands made of deoxynucleotides where nucleotides are linked between their phosphate and deoxyribose portions. In DNA, there are four types of nucleotides, which respectively contain adenine (A), guanine (G), thymine (T), and cytosine (C) bases. The two strands of a DNA molecule are held together by hydrogen bond formation between complementary base pairs. EX1031 (Garland), 3. Specifically, guanine (G) is complementary to and pairs with cytosine (C) and adenine (A) is complementary to and pairs with thymine (T). EX1031 (Garland), 3. Each nucleotide has a 5’ phosphate (“5’” or “5’-P”) end and a 3’ hydroxide (“3’” or “3’-OH”) end that links to the 5’ end of the next nucleotide. During extension of a given DNA by enzymatic or chemical means, nucleotides are added to the 3’ end of the existing chain through the catalyzed reaction of a deoxynucleotide triphosphate (dNTP) being used to add the nucleotide releasing a diphosphate in the reaction. EX1031 (Garland), 38.

45. Knowing the sequence of a given DNA can be important for any number of reasons, including disease diagnosis, identifying disease-causing mutations, and confirming genetic identity. For the last five decades, many DNA-sequencing techniques have been developed, most of which are based on the complementarity of the two strands in DNA.

46. In 1977, Dr. Frederick Sanger published his method of DNA sequencing, now known as “Sanger-based sequencing,” which uses a technique known as non-reversible termination to determine the order of nucleotides in a strand of DNA. EX1009 (Sanger 1977). The traditional Sanger-based sequencing workflow began by amplifying a sequence of interest into millions of copies through *in vivo* cloning. EX1012 (Shendure), 2. Genomic DNA of a target sequence was fragmented, cloned into plasmid vectors, and then transformed to *E. coli* for amplification. EX1012 (Shendure), 2. For each sequencing reaction, a single bacterial colony was picked and plasmid DNA is isolated. EX1012 (Shendure), 2.

47. Following amplification and isolation, reverse strand synthesis was performed on these copies using a known priming sequence upstream of the region to be sequenced. EX1010 (Kircher), 2-3; EX1011 (Metzker), 1-2; EX1012 (Shendure), 1. Reverse strand synthesis used four pools of polymerization reactions, each with a mixture of dNTPs (i.e., dATP, dGTP, dCTP, and dTTP), plus one chemically labeled dideoxynucleoside triphosphate, a dNTP missing a hydroxyl

group at the '3-end (i.e., ddATP, ddGTP, ddCTP, or ddTTP). EX1012 (Shendure), 1; EX1010 (Kircher), 2. In other words, four reactions were performed (consisting of dNTPs plus ddATP, ddGTP, ddCTP, or ddTTP), wherein polymerization of the nucleotides occurred by incorporation of dNTPs. EX1012 (Shendure), 1. Polymerization was stochastically terminated wherever the labeled dideoxynucleotides happened to be incorporated. EX1010 (Kircher), 2-3; EX1011 (Metzker), 1-2; EX1012 (Shendure), 1. Incorporation of a dideoxynucleotide into the growing strand terminated DNA polymerization because dideoxynucleotides lack the 3'-OH group necessary for the phosphodiester bond formation that would have allowed polymerization to continue. EX1010 (Kircher), 2-3; EX1011 (Metzker), 1-2; EX1012 (Shendure), 1. The dNTPs/ddNTP mixture thus causes random, non-reversible termination of the extension reaction, creating different copies of molecules extended to different lengths. EX1010 (Kircher), 2-3; EX1011 (Metzker), 1-2; EX1012 (Shendure), 1-2.

48. The resulting molecules were sorted by molecular weight via electrophoresis (corresponding to the point where a labeled dideoxynucleotide terminated polymerization), and signals generated from labels attached to the terminating dideoxynucleotides were detected to determine which nucleotides were added. EX1010 (Kircher), 2-3; EX1011 (Metzker), 1-2; EX1012 (Shendure), 1-2.

49. Although significant improvements were made to the technique after 1977, Sanger-based sequencing remained time-consuming and expensive compared to later-developed techniques (“next generation sequencing”), described below. EX1010 (Kircher), 2. Among other disadvantages, Sanger-based sequencing required isolating every molecule to be sequenced separately and thus was limited to sequencing only one target sequence in each reaction.

50. Consequently, Sanger-based sequencing was ill suited for sequencing thousands of target sequences in parallel. For example, Sanger-based sequencing was used in the first project of the Cancer Genome Atlas. *See* EX1035 (TCGA), 6. But it required massive amplification of DNA from more than one hundred specimens and then further amplification of thousands of individual targets to be sequenced in hundreds of samples. *See* EX1035 (TCGA), 1, 6.

B. Next Generation Sequencing Techniques Available as of December 2010

51. Starting in the mid-2000s, “next-generation sequencing” techniques became commercially available, which greatly increased efficiency compared to traditional Sanger-based sequencing. *See* EX1022 (Li), Abstract. Among other advantages, “next generation sequencing” techniques facilitated first parallel sequencing (sequencing millions of molecules under one set of reaction conditions) and later “multiplex sequencing”—that is, the simultaneous sequencing of templates

from multiple samples during a single run. *See infra* § V.B.2; EX1023 (Choi), 1 (reporting “massively parallel DNA sequencing” using next-generation sequencing).

52. As of December 2010, Illumina and Roche 454 were two of the major “next-generation” sequencing platforms known in the art. *See* EX1004 (Parameswaran), 1. Here, I provide a brief overview of the basic principles and the sequencing workflow behind Illumina and Roche 454 sequencing.

1. Overview of Sequencing Workflow

53. Shortly after launching, “next-generation sequencing” techniques such as Roche 454 and Illumina “outperform[ed] the older Sanger-sequencing technologies by a factor of 100-1,000” and simultaneously “reduc[ed] the cost of sequencing.” EX1010 (Kircher), 1, 3, 5, 11. While the sequencing biochemistries used for Roche 454 and Illumina platforms differ, “their work flows are conceptually similar.” EX1012 (Shendure), 2.

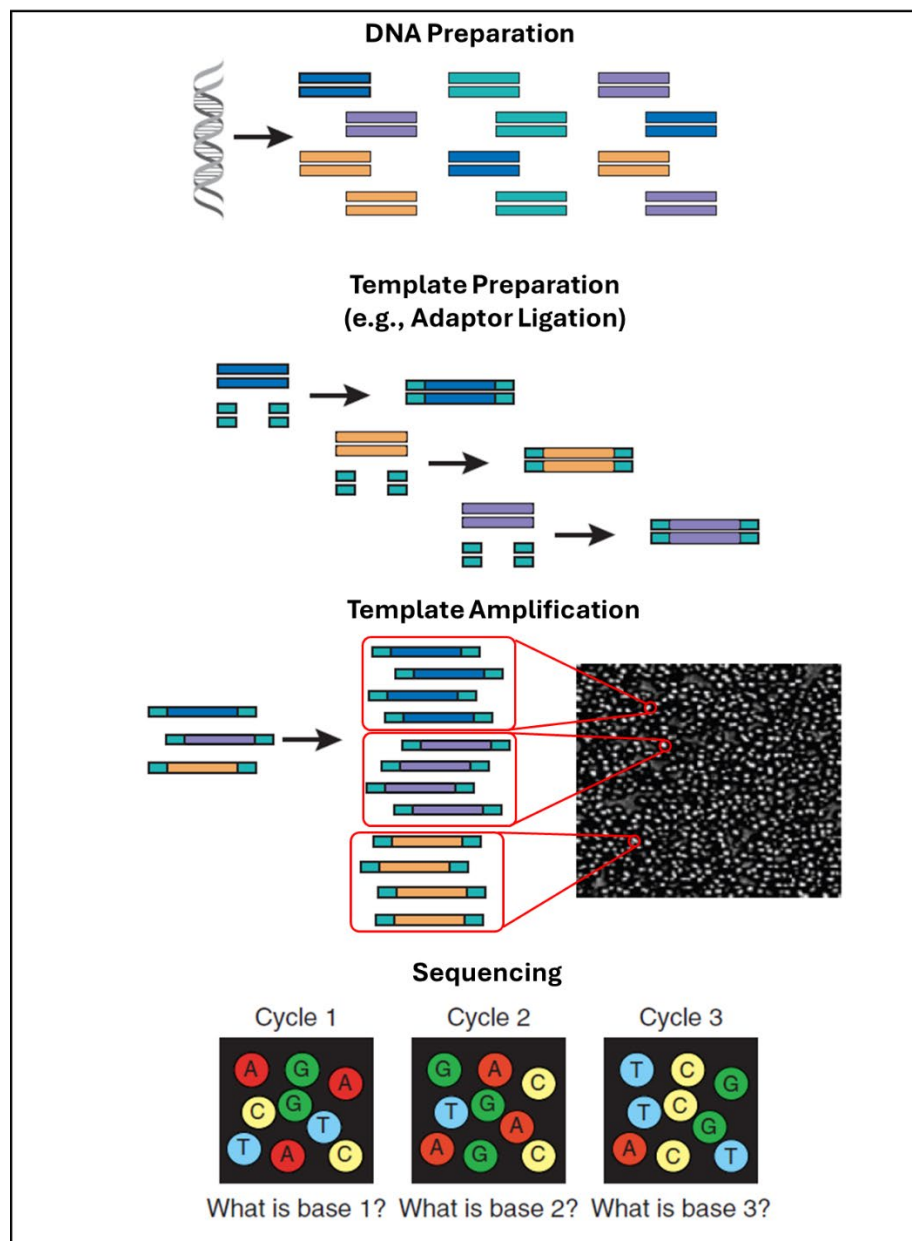


Figure 1. General Sequencing Workflow
(Revised from EX1012 (Shendure), 2 (Figure 1(b)))

54. As shown in **Figure 1**, both platforms start with sample preparation. This involved extracting DNA from a sample (or extracting RNA and generating cDNA). EX1012 (Shendure), 2 (Figure 1b). Sample preparation also involved

attaching sequencer-specific adapters, which were used to immobilize copies of the DNA on a solid substrate during sequencing. Both platforms commonly used PCR to amplify the adapter-bound DNA molecules prior to sequencing. EX1012 (Shendure), 2 (Figure 1b); EX1010 (Kircher), 4. While Roche 454 and Illumina platforms used their own platform-specific adapters, they shared many principles of operation and configuration. EX1021 (Craig), 1-2, 15; EX1014 (Meyer 2008), 2. Specifically, both platforms spatially amplify DNA on a substrate to run alternating cycles of enzyme-driven biochemical reaction and imaging-based data acquisition. EX1012 (Shendure), 2. In the following sections, each step is discussed in detail.

2. Basic Principles Behind Illumina and Roche 454 Sequencing

55. Unlike the traditional Sanger-based approach which sequenced only one target sequence per reaction, the Illumina and Roche 454 sequencing platforms provided higher throughput, resulting in more reads and substantial cost savings over traditional Sanger-based sequencing. For example, as of December 2010, it was estimated that the traditional Sanger-based method cost about \$500 per megabase of sequence data. For the same amount of sequence data, Roche 454 sequencing cost about \$20 and Illumina sequencing cost about \$0.50. *See* EX1010 (Kircher), 10 (comparing throughput, read length, sequencing quality, and cost of Sanger-based sequencing, Roche 454 sequencing, and Illumina sequencing).

56. Illumina and Roche 454 sequencing platforms both use versions of a method known as sequencing-by-synthesis, by which the sequences of single-stranded template nucleic acids are determined by adding specific bases to the nucleic acid strands and then detecting the nucleotides incorporated into that strand (i.e., A, C, G, or T). EX1010 (Kircher), 3-6. While both Illumina and Roche 454 sequencing use sequencing-by-synthesis, different signals are used to detect the nucleotides incorporated into the complementary strand. EX1011 (Metzker), 4-5, 8-9; EX1010 (Kircher), 3-6.

57. **Figure 2** below provides a schematic drawing of Illumina sequencing, which became widely available by 2007. Illumina sequencing detects a unique fluorescent signal incorporated into the newly synthesized strand during each sequencing round, and this signal shows what nucleotide was incorporated. EX1010 (Kircher), 5. Specifically, Illumina sequencing uses fluorescently labeled deoxynucleotide triphosphates (dNTPs) bound to reversible terminators. EX1006 (Bentley), 1; EX1011 (Metzker), 4-5; EX1010 (Kircher), 4-6. After incorporation of each nucleotide, the polymerization reaction is halted because of the reversible terminator; no additional base can be added. EX1010 (Kircher), 4-6; EX1011 (Metzker), 4-5. The polymerase and free dNTPs are washed off, and the fluorescent signal associated with the incorporated nucleotide is detected. EX1010 (Kircher), 4-6; EX1011 (Metzker), 4-5. The fluorescent dye and terminator are then

cleaved, the 3'-OH group necessary for phosphodiester bond formation is regenerated, and the polymerization reaction is reconstituted with enzyme and dNTPs. EX1010 (Kircher), 4-6; EX1011 (Metzker), 4-5. **Figure 2** depicts the time lapsed fluorescent signal readout for each respectively incorporated nucleotide.

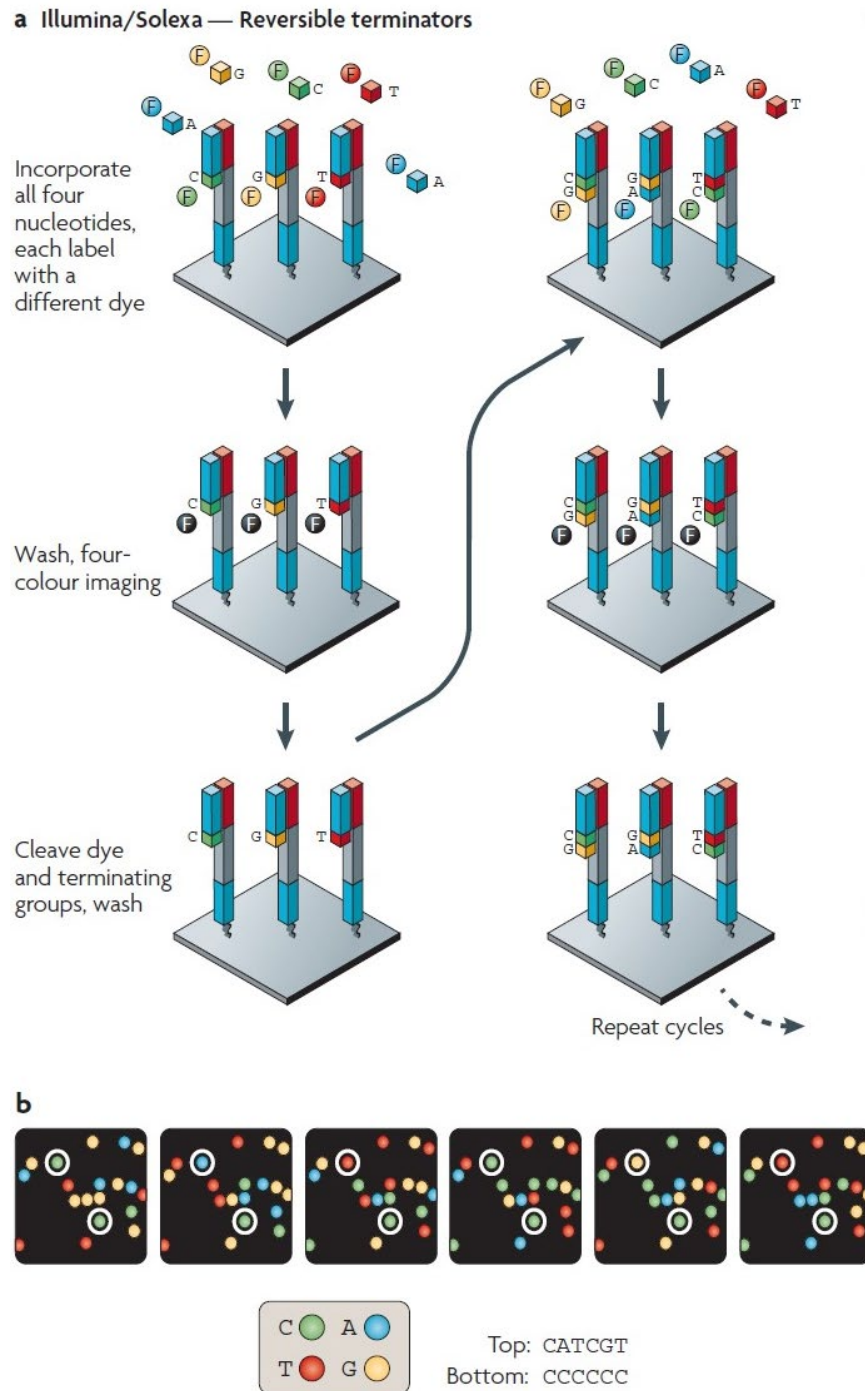


Figure 2. Illumina Sequencing
(EX1011 (Metzker), 6 (Figure 2))

58. Roche 454 sequencing was first released in October 2005. Roche 454 sequencing uses a variant of sequencing by synthesis known as pyrosequencing,

which detects which nucleotides are incorporated based on the release of pyrophosphate and consequent generation of photons. EX1011 (Metzker), 8-9; EX1010 (Kircher), 3-4. During pyrosequencing, each of the four nucleotides sequentially flows over the template. EX1011 (Metzker), 8-9; EX1010 (Kircher), 3-4. If the added nucleotides are complementary to the template, polymerases will incorporate the nucleotides. EX1011 (Metzker), 8-9; EX1010 (Kircher), 3-4. This incorporation event produces a pyrophosphate, which is converted to a detectable light signal. DNA chains complementary to the template are extended by polymerase when the next nucleotide needed to extend the chain is present in the reaction and detected as light. EX1011 (Metzker), 8-9; EX1010 (Kircher), 3-4. The remaining unincorporated nucleotides are washed away, and the next nucleotide is flowed over the template until all four nucleotides have been flowed. EX1011 (Metzker), 8-9; EX1010 (Kircher), 3-4. This cycle can be repeated until the full complement to the template is synthesized. EX1010 (Kircher), 3-4; EX1011 (Metzker), 8.

3. Immobilization of Nucleic Acids of Interest

59. Both Illumina and Roche 454 sequencing use a “flow cell” where fluids containing sequencing reagents can flow over the templates and their complementary strands in a controlled manner. EX1010 (Kircher), 3-4. In Illumina and Roche 454 sequencing, both ends of the template nucleic acid contain platform-

specific adapter sequences that anneal (attach by hybridization) the template to complementary nucleotide sequences on the surface of substrates to keep those templates in place. EX1010 (Kircher), 3-6; EX1012 (Shendure), 2-6.

60. For Illumina sequencing, that substrate is a material coated on the top of the flow cell, which contains oligonucleotides that are complementary to the Illumina adapter sequences. EX1010 (Kircher), 4-6; EX1025 (Illumina Technology Spotlight), 1-2; EX1011 (Metzker), 2-3; EX1012 (Shendure), 4-6. For Roche 454 sequencing, the substrate is a microbead coated with oligonucleotides complementary to the Roche 454 adapter sequences. EX1010 (Kircher), 3-4; EX1011 (Metzker), 2-3; EX1012 (Shendure), 3-4. Nucleic acid-bound beads are then placed into Roche 454 miniaturized wells, which are arranged as a plate that houses one substrate coated bead per well. EX1010 (Kircher), 3-4; EX1011 (Metzker), 2-3, 8-9; EX1012 (Shendure), 3-4.

4. Amplification of Nucleic Acids of Interest

61. For sequencing by synthesis methods, additional amplification of template molecules is typically needed to generate detectable signals during sequencing. As such, the extracted nucleic acids must be copied multiple times (i.e., amplified) such that the signals generated during sequencing achieve detectable levels. EX1011 (Metzker), 2.

62. Amplification relies on a laboratory process called the polymerase chain reaction (“PCR”). EX1010 (Kircher), 4-5. As depicted in **Figure 3** below, during PCR, the two strands in a DNA molecule are separated, and each strand serves as a template for generating a reverse complementary strand in the presence of primers, free nucleotides, and polymerase enzymes. EX1032 (PCR Background), 1. The reaction is repeated, yielding exponential amplification such that thousands or millions of copies of the original DNA molecules are generated. Many DNA sequencing techniques, including both Illumina sequencing and Roche 454 sequencing, incorporate PCR into the workflow, often several times.

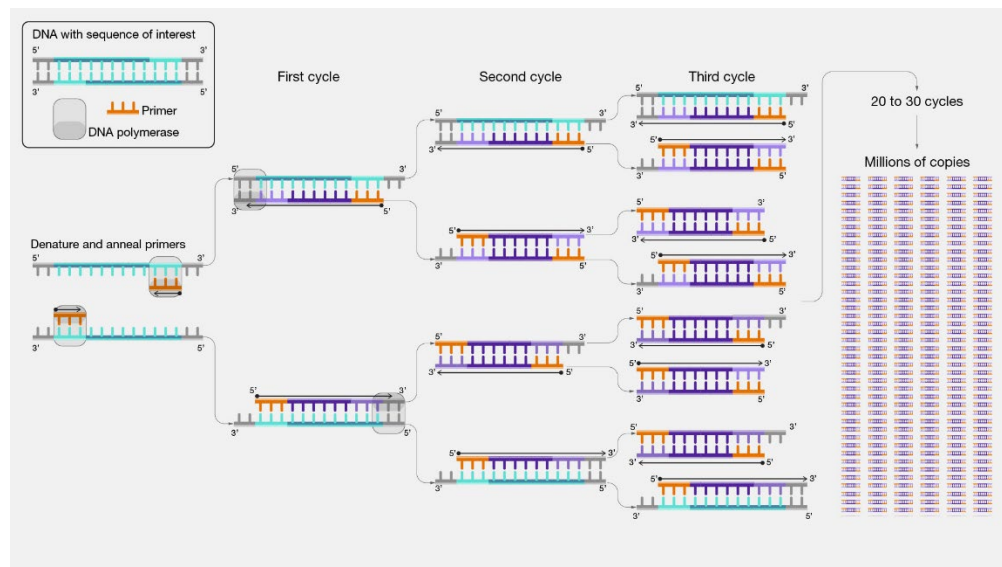


Figure 3. Schematic Drawing of PCR Reaction
(EX1032 (PCR Background), 1)

63. Both Illumina and Roche 454 sequencing require amplifying the templates to form thousands of copies prior to sequencing. EX1010 (Kircher), 4-5;

EX1012 (Shendure), 2. In Illumina sequencing, amplification occurs on the surface of a flow cell. EX1012 (Shendure), 4. Specifically, Illumina sequencing uses “bridge amplification,” depicted below in **Figure 4**, where each template is immobilized on the surface of a flow cell, folds over, and forms an arched, bridge-like shape for amplification. EX1012 (Shendure), 4. When the template is in this bridge position, a complementary strand is formed. EX1012 (Shendure), 4. Next, the resulting double-stranded molecule is denatured, leaving one template strand and one complementary strand attached to the flow cell, both of which then form new bridges to continue amplifying that template sequence. EX1012 (Shendure), 4. The result of this amplification process is a cluster of multiple copies of the template all bound to the surface of the flow cell in a narrowly defined area. EX1010 (Kircher), 5; EX1025 (Illumina Technology Spotlight), 1-2; EX1011 (Metzker), 2-3; EX1012 (Shendure), 4. This cluster of templates gives a stronger signal that is more easily read than a single strand alone would give, allowing a more accurate reading of which nucleotides are being incorporated in each sequencing step.

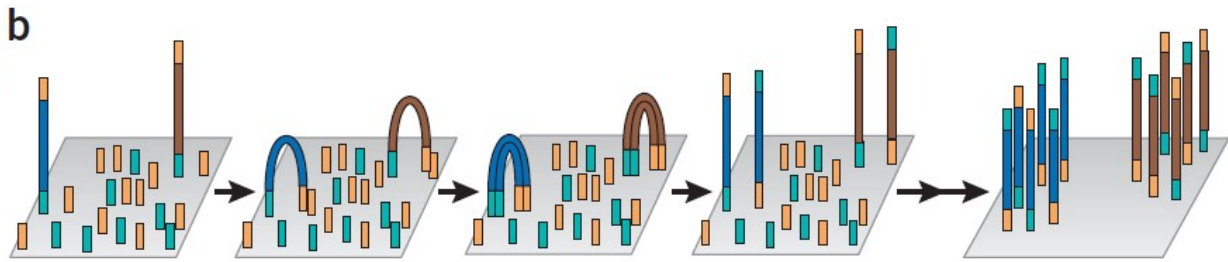


Figure 4. Bridge Amplification
(EX1012 (Shendure), 4 (Figure 2b))

64. In Roche 454 sequencing, amplification occurs on the surface of spherical solid substrates (microbeads). Specifically, Roche 454 sequencing uses “emulsion” PCR, which effectively creates millions of parallel PCRs amplifying the templates on the surface of beads. EX1010 (Kircher), 4; EX1011 (Metzker), 2-3; EX1012 (Shendure), 3-4. The result is similar to Illumina sequencing in that a large number of copies of the same template are bound to the surface of a single feature (bead), giving a stronger signal during sequencing.

5. Generation of Sequence Reads and Sequences of Nucleic Acids of Interest

65. Each sequencing run generates “sequence reads.” The length of sequence reads can vary. As of December 2010, Roche 454 protocols allowed reads of up to 200 to 300 bp nucleotides in length. EX1012 (Shendure), 4. When first introduced in 2007, Illumina sequencing reads typically were about 36 nucleotides in length without a high error rate. EX1012 (Shendure), 5. But as of July 2010, Illumina sequencing generated reads of about 150 nucleotides in length. EX1027

(Illumina Press Release), 1. Depending on the length of the template nucleic acids being sequenced, a given sequence read could therefore provide the partial or complete sequence for the corresponding template nucleic acid.

66. By December 2010, both Illumina and Roche 454 platforms could be used for paired-end sequencing, which involved generating two sequence reads for the same target sequence, one running from each end of the target sequence. EX1010 (Kircher), 5 (describing Illumina paired-end sequencing); EX1013 (Wiley) (disclosing Roche 454 paired-end sequencing); EX1027 (Illumina Press Release), 1. Paired-end sequencing provided multiple advantages, including increasing the overall length of template nucleic acids that could be sequenced. By combining reads from both ends of a template nucleic acid and matching up the two reads, one could completely sequence template nucleic acids that would otherwise be longer than the sequencer's allowable read length. The use of paired end sequencing was well established in the art, being foundational to the first draft human genome sequence. The basic concept of paired-end sequencing is depicted in **Figure 5** below.

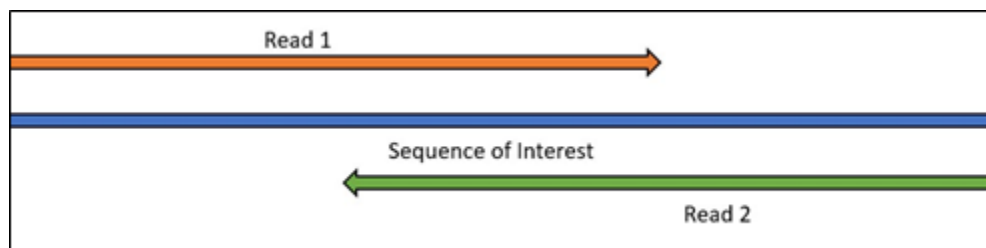


Figure 5. Paired-End Sequencing

6. Use of Sequence Identifiers in Multiplex Sequencing

67. As discussed above, next-generation sequencing platforms were advantageous because they could produce so much sequence at such little cost. By December 2010, it became standard to sequence many samples in parallel, referred to as “multiplex” sequencing. *Supra* § V.B. Multiplex sequencing relies on the traceability of nucleic acids (i.e., the ability to determine the source of a given nucleic acid after its sequence is determined). This involved uniquely labeling nucleic acids derived from multiple sources so that the sequences derived from the labeled nucleic acids can be assigned to the correct source.

68. Well before December 2010, researchers performed multiplex sequencing by incorporating sample-specific “identifier sequences” into the sequencing workflow. In this process, sample-specific “identifier sequences” were incorporated into nucleic acids from each sample, which were then sequenced simultaneously on the Illumina or Roche 454 platforms. The resulting sequencing reads, which determined the sequence of both the identifier sequence(s) and the template sequence, were then “demultiplexed” (i.e., allocated to the correct source) according to the identifier sequences with which they were associated. Sequences with missing, truncated, or altered identifier sequences were discarded. In the context of DNA sequencing, the term “identifier sequence” is synonymous with the terms “barcode,” “index,” and “tag.”

69. When next-generation sequencing methods were first introduced, researchers used one identifier sequence per template (i.e., “single indexing”). E.g., EX1015 (Meyer 2010), 4; EX1016 (Toloue), [0089]-[0090]. But researchers soon noticed limitations with single-indexing methods. For instance, single-indexing methods were more susceptible to errors resulting from primer synthesis, primer mismatch, damage to the original template, or errors during the sequencing process that mis-reported or disconnected the two ends of the sequence. *See* EX1004 (Parameswaran), 7; EX1029 (Binladen), 6. Using single-indexing methods, it could be difficult to distinguish sequence reads that were assigned (or contained) the correct indexes from those with incorrectly assigned (or contained) indexes, and thus template sequences were sometimes misassigned to the wrong source.

70. Problems with single indexes could also arise during PCR amplification when chimeric molecules of the template nucleic acids were formed—that is, erroneous combinations of at least two individual nucleic acid molecules that are from a single sample or from two different samples. EX1017 (Lahr), 1 (“Chimeras are formed when incompletely extended DNA fragments anneal to closely related sequences generating recombinants between starting templates.”); EX1018 (Meyerhans), 1 (“[R]ecombinant or ‘shuffled’ molecules, i.e. molecules composed of parts of two different sequences artificially combined during PCR may be produced.”); EX1019 (Paabo), 1 (“[L]esions such as breaks, apurinic sites, and UV

damages in a DNA template can cause the extending primer to jump to another template during the polymerase chain reaction.”); EX1020 (Odelberg), 7 (“a template-switch can occur between two nascent extending strands.”). **Figure 6** below depicts a chimeric sequence resulting from the combination of template #1 and template #2.

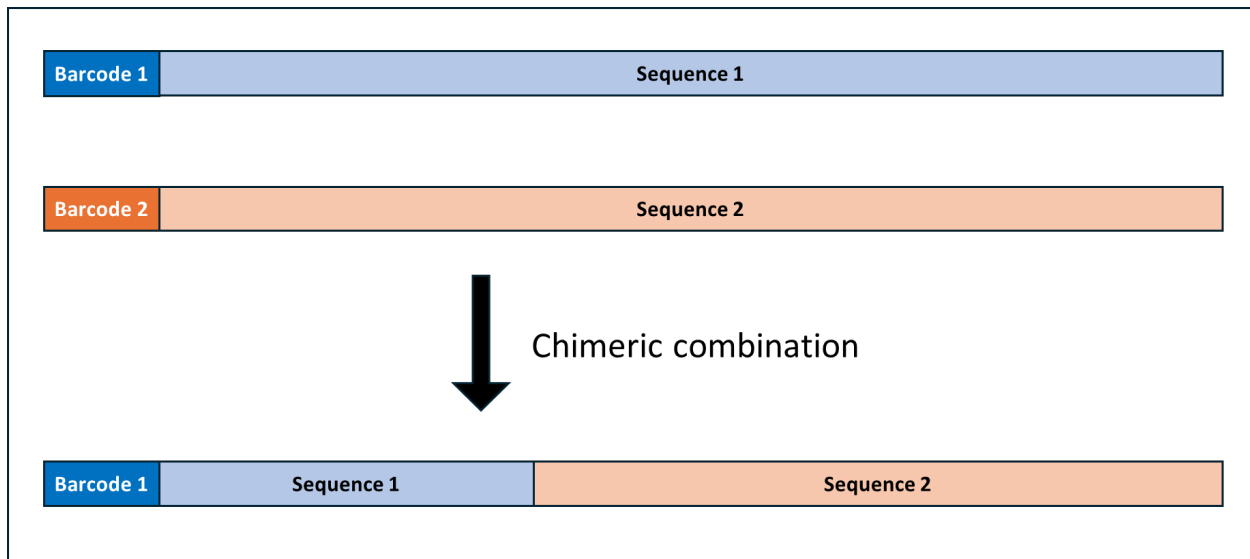


Figure 6. Formulation of Chimeric Molecule During Sequencing Workflow

71. As of December 2010, chimeric sequences generated during sequencing workflow were a widely recognized problem for next-generation sequencing platforms, including both Roche 454 and Illumina sequencing. *See, e.g.*, EX1034 (Creer), 10, 13 (characterizing next-generation sequencing as “the ideal ‘breeding ground’ for [chimeric] DNA molecules,” especially when the experiments involve “the amplification of homologous regions from a large number of potentially highly related organisms”); EX1033 (Dai), 8-9 (reporting chimeric

sequence reads from Illumina sequencing). The presence of chimeric sequences was “worrisome” because “[i]t can be difficult to differentiate original haplotypes from chimeras” (i.e., differentiate sequences that were made of two independent molecules from those without an error). EX1017 (Lahr), 1; *see also* EX1033 (Dai), 8-9 (“A key issue in paired-end/mate pair data analysis is the false chimeric sequences” from “chimeric sequences [that] are in fact candidates for gene fusion, inversion, insertion and/or deletion.”).

72. Chimeric molecules arising from the combination of templates from different samples were problematic for multiplex sequencing, leading to misassignment of sequences to incorrect samples and lowering the overall accuracy of sequencing. As illustrated in **Figure 6** above, when only one identifier sequence is used for sample identification, sequences resulting from chimeric molecules can be incorrectly assigned to different samples during demultiplexing. *See* EX1017 (Lahr), 1. And, because such chimeric sequences may still have a barcode sequence that matches one of the barcodes being used in the multiplex reaction, it may not be possible to determine that the sequence is a chimeric sequence.

73. By December 2010, it was also appreciated that “dual-indexing” methods—that is, methods in which *two* identifier sequences, one at each end of the molecules in a sample, were assigned to each sample—better facilitated identification and removal of incorrect sequences arising from the technical errors

discussed above. *See* EX1004 (Parameswaran), 7; EX1029 (Binladen), 5; EX1038 (Maughan), Abstract (using “multiplex identifier (MID) barcodes” to sequence “pooled genomic reduction libraries” in “next-generation sequencing (454-pyrosequencing and Illumina GAlle)”), 4 (“barcode primer pairs” were attached to “DNA fragments for each DNA sample”); *supra* ¶ 70. As explained below, at least two forms of dual-indexing were known as of December 2010: unique dual-indexing and non-unique dual-indexing.

74. Unique dual-indexing uses two identifier sequences per sample and does not permit an identifier sequence to be used with more than one sample. *See, e.g.*, EX1038 (Maughan), 3 (Table 1). Template nucleic acids from the same sample are therefore associated with a “distinct” or “unique” pair of identifier sequences, and each identifier sequence can be used only once across all samples. *Parameswaran*, for example, describes a unique dual-indexing method. EX1004 (Parameswaran), 2 (“[W]e describe a high-information-content barcoding approach in which each sample is associated with two uniquely designed, 10-nucleotide barcodes.”), 14-15 (Supplementary Table 1A reciting identifier sequences that never repeat across samples).

75. **Figure 7** below provides an illustration of unique dual indexing. Here, identifier sequences at the 5’ end of the templates are represented by letters (A, B, C, etc.) and identifier sequences at the 3’ end of the templates are represented by

numbers (1, 2, 3, etc.). In addition to all combinations of identifier sequences (e.g., “A-1” or “I-9”) being unique to specific samples, each individual identifier sequence (e.g., A, I, 1, 9, etc.) is also unique to a specific sample. For instance, the 5’ index of “A” is used in combination with the 3’ index of “1,” forming the identifier pair A-1, and neither “A” nor “1” is repeated for any other sample.

	5’ Index (A, B, C, etc.)					
3’ Index (1, 2, 3, etc.)	A-1	B-2	C-3	D-4	E-5	F-6
	G-7	H-8	I-9	J-10	K-11	L-12
	M-13	N-14	O-15	P-16	Q-17	R-18
	S-19	T-20	U-21	V-22	W-23	X-24

Figure 7. Unique Dual-indexing

76. Because unique dual-indexing assigns templates from each sample to a unique pair of identifier sequences, sequences that contain a pair of identifier sequences that was not assigned to any sample reflect chimeric sequences, including those derived from cross-over errors, and thus can be discarded. EX1004 (Parameswaran), 7; EX1029 (Binladen), 6. For example, in the table above, a sequence containing an expected pair of a 5’ index of “A” and index of “1” (i.e., A-1) would be kept, but a sequence with an unexpected pair of a 5’ index of “A” and a 3’ index of “7” (i.e., A-7) could be discarded as it potentially resulted from an erroneous cross-over event or other problem during sequencing.

77. By contrast, non-unique dual-indexing also uses a pair of identifier sequences per sample, but an individual identifier sequence can be reused across different samples, so long as any given *pair* of two identifier sequences is never reused. In other words, each sample has a “distinct” or “unique” pair of identifier sequences, but the individual identifier sequences are not themselves necessarily unique to each sample. The *Gloor* reference, for example, describes a non-unique dual-indexing method. EX1005 (Gloor), 2 (“The combination of sequence tags at each end of the overlapped reads allowed us to use a small number of primers to uniquely tag a large number of samples.”); *see also* EX1016 (Toloue), [0091]-[0092].

78. **Figure 8** below provides an illustration of non-unique dual-indexing. As with the illustration of unique dual-indexing above, identifier sequences at the 5’ end are represented by letters and identifier sequences at the 3’ end are represented by numbers. Every combination of identifier sequences (A-1, B-1, A-2, B-2, etc.) is unique, but individual identifier sequences have been reused. For example, the 5’ index of “A” was reused for six samples (A-1 through A-6), and the 3’ index of “4” was reused for six samples (A-4 through F-4).

		5' Index					
		A	B	C	D	E	F
3' Index	1	A-1	B-1	C-1	D-1	E-1	F-1
	2	A-2	B-2	C-2	D-2	E-2	F-2
	3	A-3	B-3	C-3	D-3	E-3	F-3
	4	A-4	B-4	C-4	D-4	E-4	F-4
	5	A-5	B-5	C-5	D-5	E-5	F-5
	6	A-6	B-6	C-6	D-6	E-6	F-6

Figure 8. Non-Unique Dual-indexing

79. Compared to unique dual-indexing, non-unique dual-indexing may require additional steps to identify and discard chimeric sequences. Consider the same example of a sequence containing a pair of a 5' index of "A" and index of "1." This sequence could be derived from a template nucleic acid containing a correct pair of a 5' index of "A" and index of "1." But this could also result from a cross-over error where two templates, the first with a 5' index of "A" and 3' index of "2" (A-2) and the second containing a pair of a 5' index of "B" and index of "1" (B-1) are combined during amplification. This error could result in the first template sequence erroneously having an identifier pair of A-1 and thus being matched to the wrong sample. To discern that this sequence is a true chimeric sequence, additional

analyses may be needed, including comparing the sequence against a reference sequence (if known) or other sequences to determine whether it is a composite of two correct sequences. EX1005 (Gloor), 8; EX1017 (Lahr), 4-5.

80. Non-unique dual-indexing can nevertheless be beneficial over single indexing because it allows quick identification and discarding of sequences resulting from other types of technical errors, such as imperfect primer synthesis or amplification reactions.

7. Resources available to a POSA before December 2010

81. By December 2010, a POSA had access to a number of resources that made high throughput sequencing affordable and less burdensome. For example, kits for extracting nucleic acids from samples were commercially available. While these kits often came with detailed instructions, textbooks and review articles summarizing necessary laboratory techniques were also available. In addition, laboratory tools for collecting extracted nucleic acids, such as gels, columns, and beads, were commercially available.

82. Sequencer manufacturers also offered commercial kits for library preparation, i.e., generating template nucleic acids suitable for sequencing. They also offered technical support lines that would assist users in designing appropriate protocols for library preparation to meet the users' needs. For custom designed library preparation methods, commercial sources were available that would

synthesize the necessary custom oligonucleotides. *See, e.g.*, EX1038 (Maughan), 4 (purchasing barcoded primers from a commercial manufacture, Integrated DNA Technologies).

83. Some laboratories owned sequencers themselves. But many universities also had centralized laboratories that would provide sequencing services. *See, e.g.*, EX1038 (Maughan), 4 (performing Roche 454 and Illumina sequencing at the Brigham Young University DNA Sequencing Center and the HCI Microarray Core Facility at the University of Utah, respectively); EX1005 (Gloor), 23-24 (performing Illumina sequencing at the Next-Generation Sequencing Facility in the Center for Applied Genomics at the Hospital for Sick Children in Toronto). Large genome centers supported by, for example, the National Institute of Health (NIH), also housed many sequencers. In most cases, these centralized sequencing centers performed sequencing, i.e., users simply needed to drop off library preparation for sequencing. These sequencing centers also typically offered assistance for library preparation, if requested. *See, e.g.*, EX1038 (Maughan), 4; EX1005 (Gloor), 23-24.

VI. SUMMARY OF THE PRIOR ART

A. *Parameswaran* (EX1004)

84. *Parameswaran* (EX1004) is an article published in the journal Nucleic Acids Research. The article is titled “A pyrosequencing-tailored nucleotide barcode

design unveils opportunities for large-scale sample multiplexing.” EX1004 (Parameswaran), Title. On top of the first page, the article indicates that it was published online on October 11, 2007. I understand from counsel that *Parameswaran* is prior art to the '200 patent because it was published before December 23, 2010.

85. *Parameswaran* describes a “novel barcoding approach to allow for pooling and sequencing of DNA from independent samples, and to facilitate subsequent segregation of sequencing capacity.” EX1004 (Parameswaran), Abstract. *Parameswaran*’s barcoding approach uses “forward-reverse barcode pairs” in which each barcode is 10 nucleotides in length and is “unique with respect to at least 4 nt positions.” EX1004 (Parameswaran), 1, 2 (“[W]e describe a high-information-content barcoding approach in which each sample is associated with two uniquely designed, 10-nucleotide barcodes.”); *see also* EX1004 (Parameswaran), 14-17 (Supplementary Tables 1A and 1B providing a complete list of barcoded primers with barcodes between adapter and cloning linker sequences). These barcodes were designed “with care taken to avoid homopolymeric runs that

are particularly problematic in pyrosequencing technologies.”¹ EX1004 (Parameswaran), 4.

86. Using Roche 454 sequencing, *Parameswaran* sequenced cDNA from two independent sets of 25 barcoded libraries. EX1004 (Parameswaran), 7. These libraries were prepared from RNAs derived from different *in vitro* and *in vivo* virus-infected systems. EX1004 (Parameswaran), 7. As shown in **Figure 9**, three components were used to generate the forward and reverse primers used to construct these libraries, including: (1) Roche-specific forward and reverse adapters (“F-adapter” and “R-adapter”), (2) the forward and reverse barcodes (“F-barcode” and “R-barcode”), and (3) the forward and reverse cloning linkers (“F-Cloning-Linker” and “R-Cloning-Linker”).² EX1004 (Parameswaran), 2, 3 (“The template for the primers is: 454-Adapter:: Barcode:: Linker-Primer.”).

¹ A “homopolymeric” region refers to a stretch of nucleotides where multiple consecutive bases are the same. EX1012 (Shendure), 4. For example, sequencing the sequence TAGGGGTG contains a “homopolymeric” run of four guanine (G) bases in a row. EX1012 (Shendure), 4.

² Forward and reverse cloning linkers are used to construct the initial libraries of nucleic acids. EX1004 (Parameswaran), 2.

Supplementary Table 1A also provides the full sequences of forward and reverse primers used in *Parameswaran*. EX1004 (*Parameswaran*), Supp. Tables 1-2.

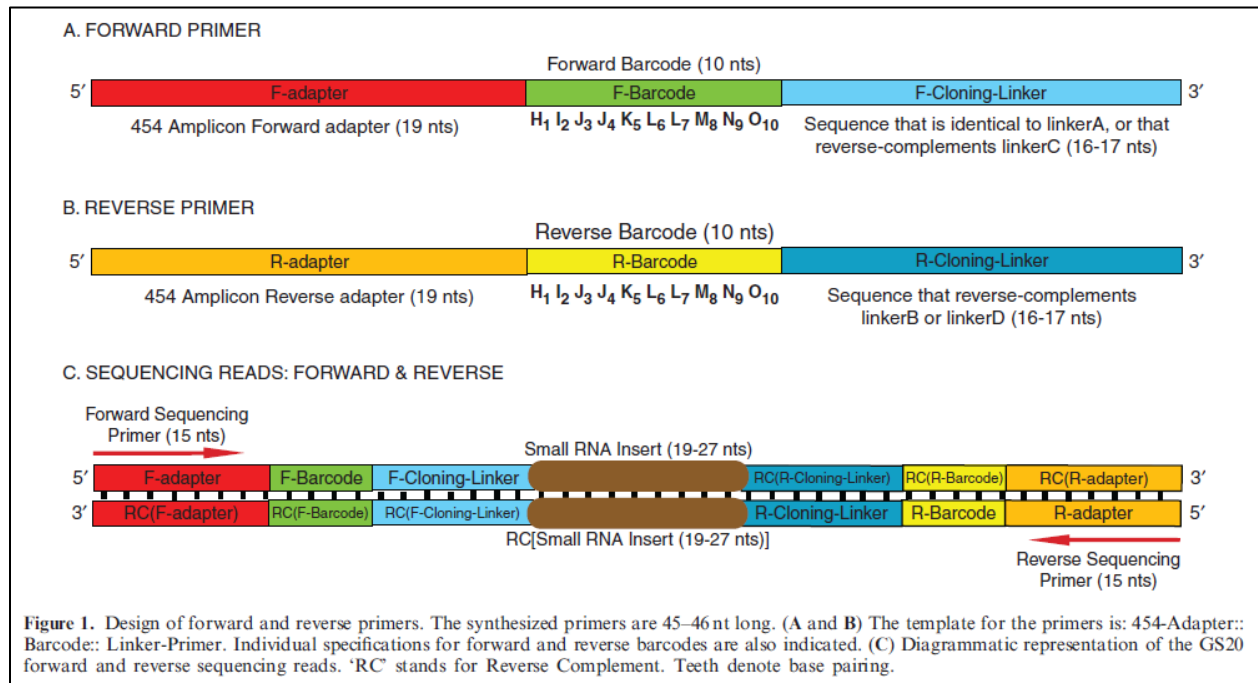


Figure 9. Design of Forward and Reverse Primers
(EX1004 (*Parameswaran*), 3 (Figure 1))

87. In two sequencing runs, *Parameswaran* obtained 237,639 sequences (run I) and 271,777 sequences (run II). EX1004 (*Parameswaran*), 7. *Parameswaran* examined the integrity of these sequences using what it calls “barcode motifs,” each of which is a 13-nucleotide combination of: (1) a 10-nucleotide barcode sequence and (2) three bases from a cloning linker. EX1004 (*Parameswaran*), 7. *Parameswaran* states that it used these 13-nucleotide barcode motifs to ensure uniqueness of the barcodes and “reduce ambiguity” in assigning sequences to the

correct source beyond what was possible with the 10-nucleotide barcode. EX1004 (Parameswaran), 7.

88. *Parameswaran* reports that “94.53% (run I) and 94.08% (run II) of the sequences had perfect forward and reverse barcode motifs, 2.90% (run I) and 3.24% (run II) of the sequences had perfect barcode motifs only at the 5’ end of the read, and 2.44% (run I) and 2.51% (run II) had perfect barcode motifs only at the 3’ end of the read.” EX1004 (Parameswaran), 7; *see also* EX1004 (Parameswaran), Table 3.

89. *Parameswaran* further divided sequences with only one perfect barcode motif match into three subcategories: “(i) perfect identity to an unexpected motif, resulting in an unassigned 5’-3’ motif combination (perfect, unexpected occurrence), (ii) a high-scoring partial match to the expected motif for that library (imperfect, expected occurrence) and (iii) no strong match to any barcode motif (imperfect, ambiguous occurrence).” EX1004 (Parameswaran), 7. In other words, sequences in the first subcategory contained an incorrect pair of identifiers, i.e., a pair comprised of one identifier correctly associated with the sample at issue and another identifier associated with another sample. Sequences in the second and third subcategories contained one correctly assigned identifier paired with an identifier that did not match or only partially matched a previously assigned identifier.

90. *Parameswaran* discusses possible technical errors that could contribute to these incorrect sequences. For example, *Parameswaran* explains that sequences in the first subcategory were attributable to “motifs with rare coincidental sequence matches (due to multiple fortuitous errors during barcode synthesis or sequencing), to contamination in PCR primer stocks, or to cross-contamination of template DNAs in the second round of PCR.” EX1004 (*Parameswaran*), 7. *Parameswaran* also explains that sequences in the second and third subcategories were attributable to “errors in sequencing,” “errors during conventional and/or emulsion PCR,” or “imperfections during the chemical synthesis of barcoded primers.” EX1004 (*Parameswaran*), 7.

91. *Parameswaran* categorized sequences in the first subcategory as a “false-discovery,” representing “misassignment” of sequences. *See* EX1004 (*Parameswaran*), 7-8 (reporting less than 0.005% of a “false-discovery rate,” which translated to a “remarkably low misassignment rate”). According to *Parameswaran*, a “remarkably low misassignment rate (0.00042% and 0.0044% for two independent experiments) was obtained by requiring a match to the 10-base barcodes and additional three bases of the fixed linker.” EX1004 (*Parameswaran*), 8.

92. *Parameswaran* states that its barcoding approach may “be compatible with other sequencing platforms.” EX1004 (*Parameswaran*), 8; *see also* EX1004 (*Parameswaran*), 2 (“We also speculate on the use of these pyrosequencer-tailored

barcodes with other sequencing platforms.”). *Parameswaran* identifies Illumina sequencing is one of the “major high-throughput sequencing platforms . . . in use today.” EX1004 (Parameswaran), 1.

B. *Gloor* (EX1005)

93. *Gloor* is an article by Dr. Gregory Gloor et al. available from arXiv.org, dated July 28, 2010. EX1005 (Gloor), 1. The article was subsequently published in the journal PLoS ONE in October 2010 with minor changes. EX1024 (Gloor PLoS ONE). *Gloor* is titled “Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products.” EX1005 (Gloor), Title. I understand from counsel that *Gloor* is prior art to the ’200 patent because it was published before December 23, 2010.

94. *Gloor* describes a “[c]ombinatorial sequence tagging” approach used to “examine hundreds of samples with far fewer primers than is required when sequence tags are incorporated at only a single end,” resulting in “a low-cost, high-throughput microbiome profiling method.” EX1005 (Gloor), Abstract. *Gloor* states that a “combination of sequence” tags was used, with one sequence tag attached to the end of each PCR primer to “uniquely” tag samples with a combination of identifiers for each sample. EX1005 (Gloor), 1-2. Table 2 of *Gloor* provides the sequences of the identifier sequence tags used, which are between three and six nucleotides long. EX1005 (Gloor), 6.

95. *Gloor* used a single “paired-end sequencing run” with Illumina sequencing to sequence the microbial V6 regions of 272 clinical samples. EX1005 (*Gloor*), 1-2. The sequencing templates described in *Gloor* contain an L-tag (left tag), a left primer, a target polynucleotide sequence, a right primer, and an R-tag (right tag) in that order. EX1005 (*Gloor*), 2-3, 6. In other words, the sequencing templates each contained the sample nucleic acid region flanked by a pair of 3-6 nucleotide-long sequence identifiers, one at each end of the template. EX1005 (*Gloor*), 1-3, 5-6 (Table 2).

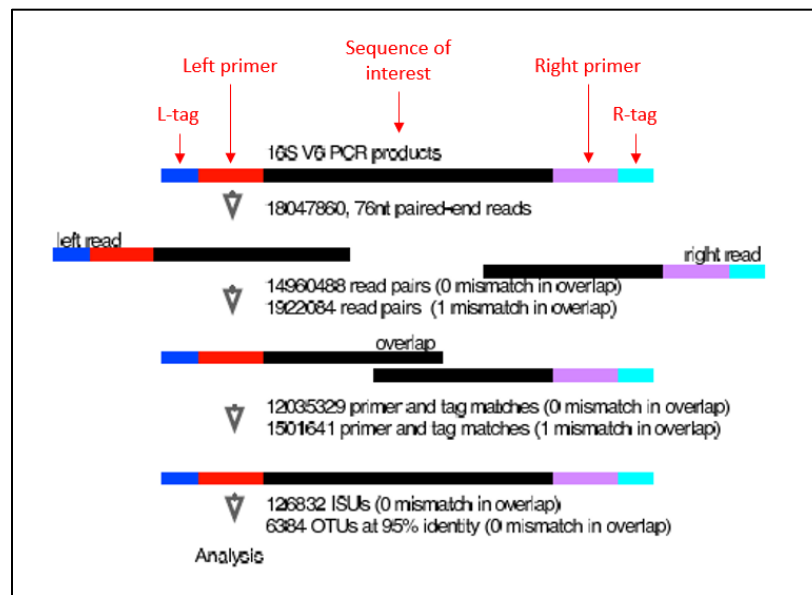


Figure 10. Paired-End Sequencing
(EX1005 (*Gloor*), 7 (Figure 2) (annotations added in red))

96. Because *Gloor* used Illumina paired-end sequencing (EX1005 (*Gloor*), Abstract), the templates also included additional standard features necessary for that platform. See EX1010 (Kircher), 4. For example, the templates

included the sequencing primer priming sites (Seq1 and Seq2) flanking the L-tag and R-tag identifier sequences, followed by the Illumina-specific adapter sequences (P5 and P7). *See* EX1010 (Kircher), 4.

97. Of note, *Gloor* states that the sequence tags were incorporated at the “3” end of the PCR primers. EX1005 (*Gloor*), 2, 6, 22. A POSA would have understood that this was an error in the context of the primer designs discussed in *Gloor*. As shown in **Figure 10** above, a POSA would have understood that the blue tags on the left and right sides of the PCR amplicon were at the 5’ end of the left and right primers (not the 3’ end). In contrast, the 3’ end of the left and right PCR primers are complementary to the 967-985 and 1078-1061 coordinates of the 16S mRNA segments, respectively. EX1005 (*Gloor*), 3. The fact that *Gloor*’s statement was erroneous is confirmed by its correction in a later version of *Gloor* published in October 2010. EX1024 (*Gloor* PLoS ONE), 1, 3, 11.

98. **Figure 10** above depicts the sequencing workflow used in the study. Illumina paired-end sequencing generated two sequence reads from one template nucleic acid, one read beginning from each end of the template and each 76 nucleotides in length. EX1005 (*Gloor*), 7 (Figure 2). Because the template was about 110-130 nucleotides in length (EX1005 (*Gloor*), 3), each read contained either a first or second sequence identifier. EX1005 (*Gloor*), 6-7 (Figure 2). Together, these

sequence reads overlapped and provided the full sequence of the template nucleic acid.

99. Applying this workflow to 272 clinical samples, *Gloor* reports that about 18 million sequence reads that were 76 nucleotides in length were collected. EX1005 (*Gloor*), 6-7. Using computer software, *Gloor* identified overlapping sequence reads and extracted about 12 million sequences that “contained two valid sequence tags and both primer sequences; allowing up to 3 mismatches per primer.” EX1005 (*Gloor*), 7. *Gloor* states that “[t]he sequences derived from the perfectly overlapping reads form the basis of the remainder of the analysis.” EX1005 (*Gloor*), 7.

100. *Gloor* also evaluated the sequences “with incorrect sequence tags” that were excluded from the remainder of the analysis. *Gloor* observed that many of these excluded sequences either completely lacked the first or second sequence identifier or had identifiers with errors that arose during synthesis or PCR amplification during the sequencing workflow. EX1005 (*Gloor*), 7-8.

101. *Gloor* states that the “Illumina sequencing method generated ~ 12 million useable reads at a cost of ~ 0.03 cents per read, an approximate order of magnitude cheaper than the per-read cost of pyrosequencing.” EX1005 (*Gloor*), 2. *Gloor* states that the “Illumina sequencing method has an exceedingly low error rate and that the majority of errors arise during the PCR amplification step.” EX1005

(Gloor), 2. And *Gloor* states that, “[s]ince Illumina sequencing is now capable of ~ 100 nt long reads from each end of a DNA fragment, the methods described here can be easily adapted for paired-end sequencing of the microbial V3, V5, V6 and the eukaryotic V9 rRNA regions.” EX1005 (Gloor), 2. *Gloor* expressed no concern about applying and using dual indexes with Illumina sequencing. *See generally* EX1005 (Gloor).

C. Binladen (EX1029)

102. *Binladen* (EX1029), titled “The Use of Coded Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by [Roche] 454 Parallel Sequencing” was published in February 2007. I understand from counsel that *Binladen* is prior art to the ’200 patent because it was published before December 23, 2010.

103. *Binladen* discusses that next-generating sequencing platforms like Roche 454 greatly improved sequencing efficiency by sequencing a large number of samples simultaneously. EX1029 (Binladen), 1-2. *Binladen* explains that “pooling” PCR products from multiple samples, however, created a problem for “the subsequent identification of the sequence reads.” EX1029 (Binladen), 1-2.

104. *Binladen* reports that it overcame this problem using a dual-indexing approach “where initial PCR primers are 5’-tagged with short nucleotide sequences (tags) in such a way that a unique tagged primer combination can be applied to each

specific DNA template source.” EX1029 (Binladen), 2. *Binladen* explains that using this method, “sequences can rapidly be sorted into their original template source using the tags.” EX1029 (Binladen), 2.

105. *Binladen* prepared two sets of tags: (1) a first set containing “sixteen unique forward[] and sixteen reverse primers through the addition of 5’ dinucleotide tags” and (2) a second set consisting of “three further unique primer pairs” each containing 5’-end “tetranucleotide tails.” EX1029 (Binladen), 3. Following PCR with these tagged primers, nucleic acid analytes from each DNA sample were flanked with unique forward and reverse sequence tags. EX1029 (Binladen), 3. The amplified products were combined and sequenced using Roche 454 sequencing. EX1029 (Binladen), 2-3. **Figure 11** below shows the sequencing workflow of *Binladen*.

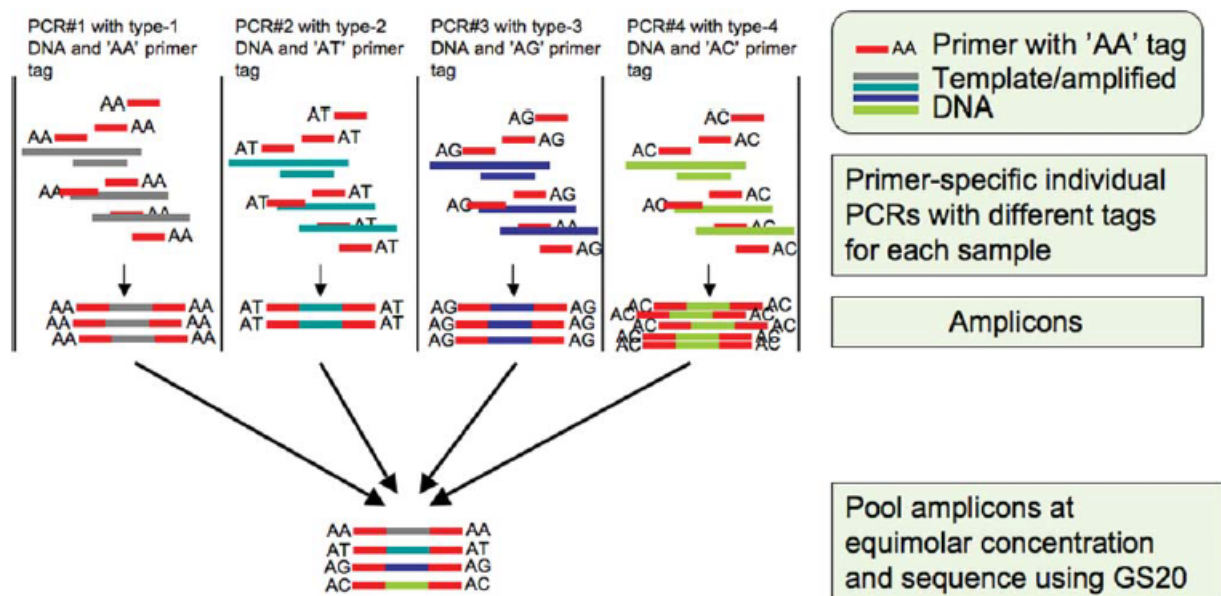


Figure 11. Sequencing Workflow in *Binladen*
(EX1029 (Binladen), 2 (Figure 1))

106. *Binladen* sequenced mammalian mitochondrial DNAs from thirteen different species using Roche 454 sequencing. EX1029 (Binladen), 3. In *Binladen*, the read lengths were about 120 bases, which were shorter than the full sequences of the amplified nucleic acid analytes (133-141 bases). EX1029 (Binladen), 3. *Binladen* explains that approximately 50% of the analytes were sequenced from the forward primer while the remaining 50% were sequenced from the reverse primer. EX1029 (Binladen), 3. *Binladen* further explains that it was therefore necessary “to label both the forward and reverse end of each PCR product.” EX1029 (Binladen), 3.

107. Because “the correct association of tags and sequences [was] crucial to the approach,” *Binladen* “discarded all sequence reads without an exact match to any of the primers used in the studies.” EX1029 (Binladen), 5. *Binladen* evaluated these discarded sequences and identified: (1) those that failed to show any match to the primer sequences in general; (2) those that matched the core primer but lacked the 5’ tag sequence; and (3) those with mismatches and/or insertion-deletions. EX1029 (Binladen), 6. *Binladen* identified incorrect primer synthesis or errors in the sequencing-by-synthesis process as possible errors that led to these discarded sequences. EX1029 (Binladen), 6.

108. *Binladen* reports that nearly 90% of the remaining sequences “diverged by no more than 1bp from one of the reference sequences, and could thus be assigned to one of the taxa analysed in the study.” EX1029 (*Binladen*), 6. *Binladen* also reports that while 4-nucleotide tag primers lowered PCR efficiency, “overall the rate of sequence miss-assignment for these primers was lower than for the dinucleotide tags.” EX1029 (*Binladen*), 7; *see also* EX1029 (*Binladen*), 8 (*Binladen* reporting “the apparent reduced miss-assignment rates of the tetranucleotide tags”). *Binladen* concludes that its approach “will be of importance to a broad range of research areas where large-scale comparisons of homologous DNA sequences from multiple sources are needed.” EX1029 (*Binladen*), 8.

D. *Bentley* (EX1006)

109. *Bentley* (EX1006), titled “Accurate whole human genome sequencing using reversible terminator chemistry” was published on November 6, 2008. I understand from counsel that *Bentley* is prior art to the ’200 patent because it was published before December 23, 2010.

110. At a high level, *Bentley* discloses Illumina sequencing, including the specific adapter sequences used during the process. More specifically, *Bentley* describes a “massively parallel synthetic sequencing approach” that first involved shearing DNA strands into double-stranded DNA fragments and ligating Illumina-specific adapters on both ends. EX1006 (*Bentley*), 2 (Figure 1a); *see also* EX1006

(Bentley), 8, 13-14. Then, the DNA fragments were denatured into single strands and annealed to complementary oligonucleotides on a coating on the flow cell. EX1006 (Bentley), 2 (Figure 1b); *see also* EX1006 (Bentley), 8, 13-14, 15.

111. The single-stranded DNA fragments were amplified on the surface of the flow cell using bridge amplification to form “clusters” of up to 1,000 identical copies of the single-stranded DNA fragments within close proximity ($\sim 1\mu\text{m}$) for sequencing-by-synthesis. EX1006 (Bentley), 2 (Figure 1b); *see also* EX1006 (Bentley), 8, 14-15. The complementary strands of the templates were then generated by re-synthesizing the second strands of the double-stranded DNAs and removing the original strands. EX1006 (Bentley), 2 (Figure 1c); *see also* EX1006 (Bentley), 8, 14-15.

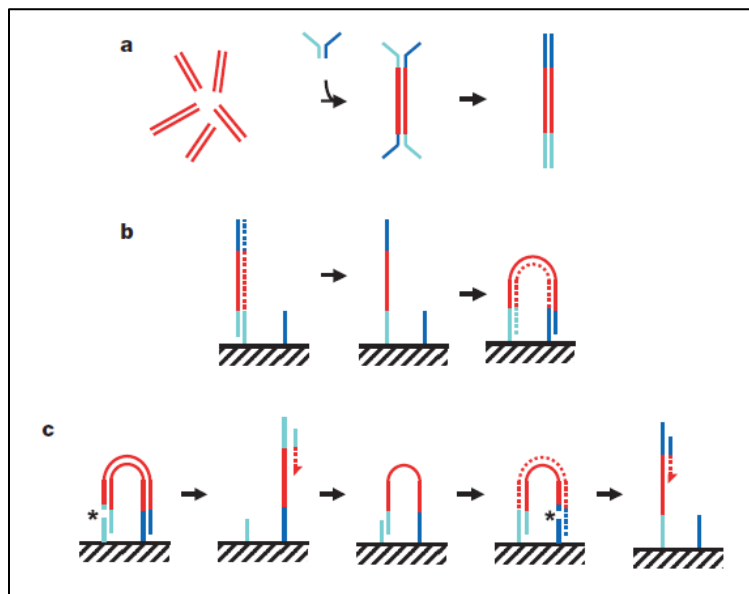


Figure 12. Preparation of Samples
(EX1006 (Bentley), 2 (Figure 1))

112. To sequence the single-stranded DNA fragments, a “set of four reversible terminators, 3’-O-azidomethyl 2’-deoxynucleoside triphosphates (A, C, G and T), each labeled with a different removable fluorophore,” was used to obtain base-by-base nucleotide incorporation in a stepwise manner. EX1006 (Bentley), 1, 26 (Figure S1a). After each cycle of incorporation, the identity of the inserted base was determined by laser-induced excitation of the fluorophores and imaging. EX1006 (Bentley), 1. The fluorescent dye and side arm were then removed to regenerate a 3’ hydroxyl group so that the next cycle of nucleotide addition could begin. EX1006 (Bentley), 1, 26 (Figure S1b).

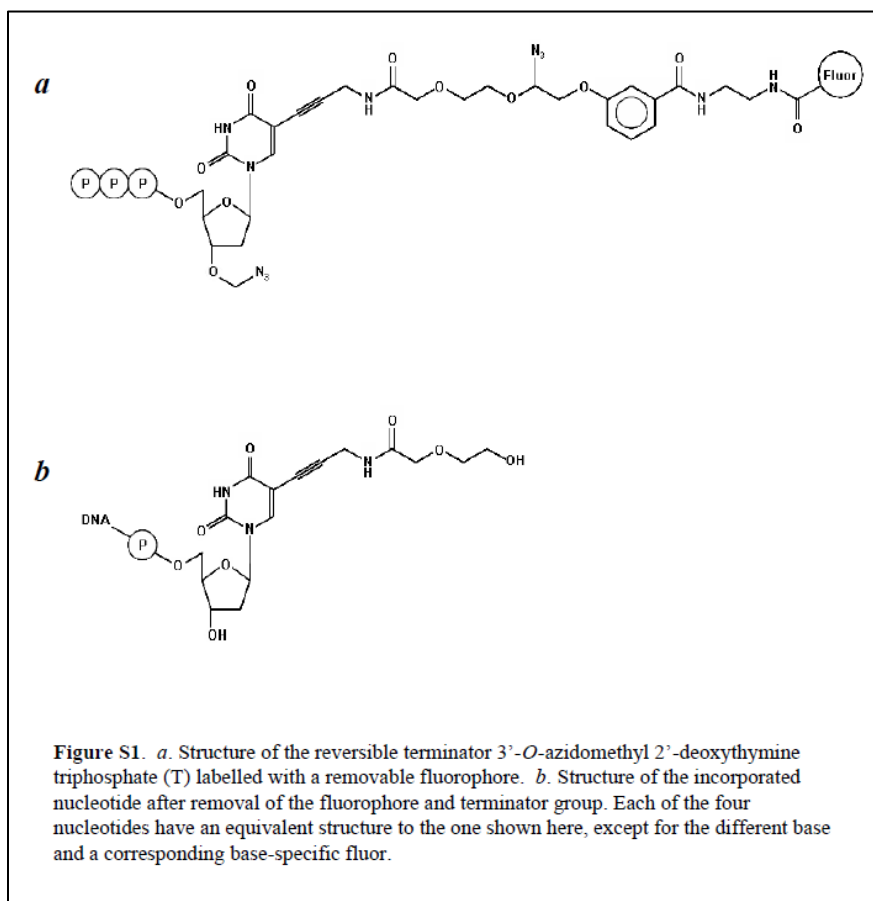


Figure 13. Structure of Reversible Termination and Incorporated Nucleotides
(EX1006 (Bentley), 26 (Figure S1))

VII. THE '200 PATENT

A. Overview

113. The '200 patent relates to “methods for maintaining the integrity and identification of a nucleic acid template in a multiplex sequencing reaction.” EX1001 ('200 patent), Abstract. The '200 patent states that, in certain embodiments, “methods of the invention involve obtaining a template nucleic acid, incorporating a pair of sequence identifiers into the template, and sequencing the template.” EX1001 ('200 patent), Abstract.

B. The Claims of the '200 Patent

114. I understand that Illumina is challenging the patentability of claims 1-6. Claim 1 is the sole independent claim.

115. Claim 1 recites:

A method for *validating the sequence* of a nucleic acid analyte of interest *in a multiplex sequencing reaction*, the method comprising:

i) *detecting the presence of* two or more identifier *sequences that are uniquely associated* with the nucleic acid analyte of interest,

wherein at least a first identifier sequence is incorporated into a 5' portion of the nucleic acid analyte of interest and wherein at least a second identifier

sequence is incorporated into a 3' portion of the nucleic acid analyte of interest, and wherein the first and second identifier sequences have four or more nucleotides and are different;

ii) **sequencing** the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence; and

iii) **validating the sequence** of the nucleic acid analyte of interest by analyzing both identifiers and **excluding the sequences** of those nucleic acid analytes of interest **containing** only one identifier or an **incorrect pair of identifiers** from sequence analysis of the multiplex sequencing reaction.

EX1001 ('200 patent), claim 1.

116. Claims 2-6, recite:

Claim 2: The method of claim 1, wherein the **identifier sequences** are **nucleic acid sequence tags**.

Claim 3: The method of claim 2, wherein the **identifier sequences** are **barcode sequences**.

Claim 4: The method of claim 3, wherein the **barcode sequences** are from **4 to 20 nucleotides**.

Claim 5: The method of claim 4, wherein the **barcode sequences** are designed to have **no homopolymer regions**.

Claim 6: The method of claim 1, wherein the nucleic acid
analyte of interest is on the *surface of a flow cell*.

EX1001 ('200 patent), claims 2-6 (emphases added).

C. Disclosures of the '200 Patent

117. The '200 patent is directed to “methods for validating the results of a molecular detection assay and for enabling the detection of errors introduced in sample preparation.” EX1001 ('200 patent), 1:66-2:1. The methods use a pair of identifiers “designed such that each set of identifiers is correlated to a particular sample (e.g., material obtained from a particular patient or group of patients).” EX1001 ('200 patent), 3:15-19.

118. The '200 patent states that “heteroduplexes, i.e., improper pairing of template strands from different samples, may form during the amplification of the multiplexed samples.” EX1001 ('200 patent), 1:51-54. Figure 1 of the '200 Patent, copied below, provides a “set of drawings depicting one example of how molecular overlap can lead to apparent high quality, but incorrect sequence data.” EX1001 ('200 patent), 4:17-20, Figure 1.

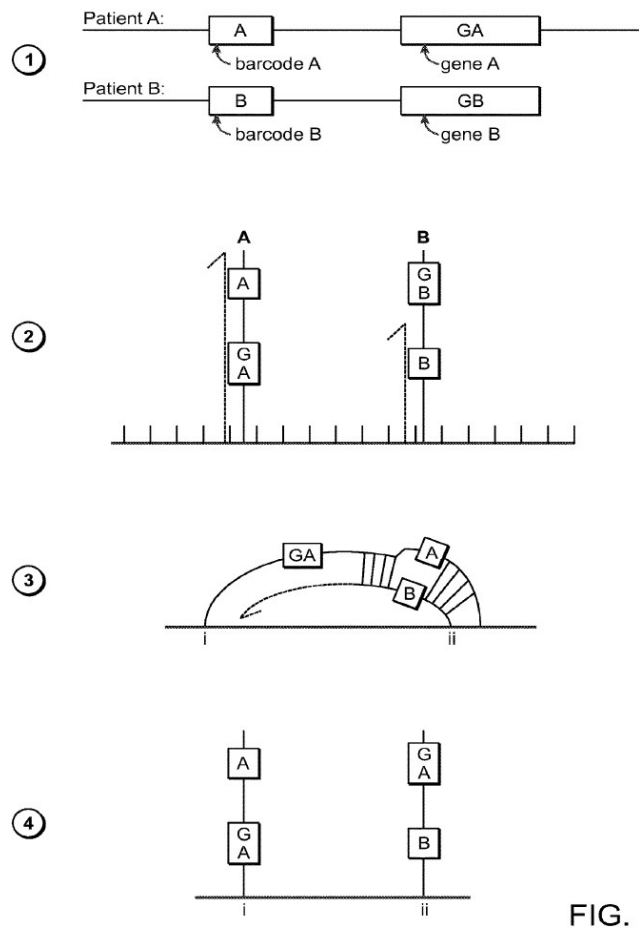


FIG. 1

Figure 14. Figure 1 of the '200 Patent (EX1001 ('200 patent), Fig. 1)

119. In **Figure 14**, which reproduces Figure 1 of the '200 patent, patient A's gene fragment (GA) is associated with one barcode (A), and patient B's gene fragment (GA) is associated with another barcode (B). EX1001 ('200 patent), Figure 1. **Figure 14** shows a heteroduplex (i.e., a chimeric molecule) formed during amplification with patient A's gene fragment and patient B's barcode. EX1001 ('200 patent), 4:40-52, Figure 1. Because only one index is used, the '200 patent

states that “patient A’s sequence data will be improperly and incorrectly associated with patient B.” EX1001 (’200 patent), 4:55-57, Figure 1.

120. The ’200 patent also states that it provides a method that uses “two or more identifiers . . . uniquely associated with an analyte to be detected.” EX1001 (’200 patent), 2:16-18. By requiring a unique combination of two or more identifiers associated with the analyte of interest, the ’200 patent states that “valid sequence data are confirmed by the presence of the pair of flanking identifiers” and “template[s] containing only one identifier or an incorrect pair (i.e., a pair or more of identifiers not associated with the same template) will be excluded from analysis.” EX1001 (’200 patent), 2:21-25.

121. Figure 2 of the ’200 Patent, reproduced below as **Figure 15**, “shows a set of drawings depicting how methods of the invention allow one to detect, during analysis, whether the integrity of a patient’s sample was maintained during a multiplex sequencing reaction.” EX1001 (’200 patent), 4:64-5:1, Figure 2.

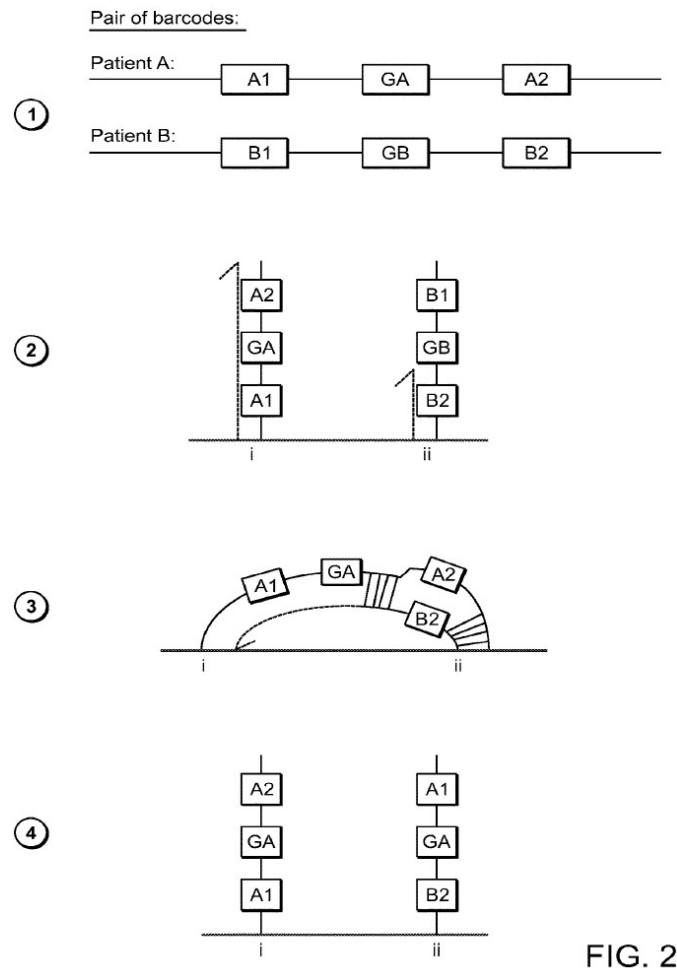


Figure 15. Figure 2 of the '200 Patent
(EX1001 ('200 patent), Figure 2)

122. In **Figure 15**, patient A's gene fragment (GA) is flanked with two barcodes (A1 and A2), and patient B's gene fragment (GB) is flanked with two other barcodes (B1 and B2). EX1001 ('200 patent), Figure 2. **Figure 15** shows a heteroduplex (i.e., chimeric molecule) formed during amplification by a cross-over error, illustrated by the fact that patient A's gene fragment is flanked by patient A's first barcode (A1) and patient B's second barcode (B2). EX1001 ('200 patent),

5:59-62, Figure 2. The '200 patent states that, “[w]hile each barcode sequence on its own is a valid barcode sequence, taken together, the two barcode sequences do not form a valid barcode sequence pair.” EX1001 ('200 patent), 5:62-65. Because a unique combination of barcodes is associated with each sample, the '200 patent states that the heteroduplex “would be recognized as invalid” during data analysis “because the data contains an invalid barcode sequence pair.” EX1001 ('200 patent), 5:66-6:1.

123. **Figure 16**, which reproduces Figure 3 of the '200 Patent, provides a “graph showing that methods of the invention dramatically reduce false positives in a multiplex sequencing reaction.” EX1001 ('200 patent), 6:7-9, Figure 3.

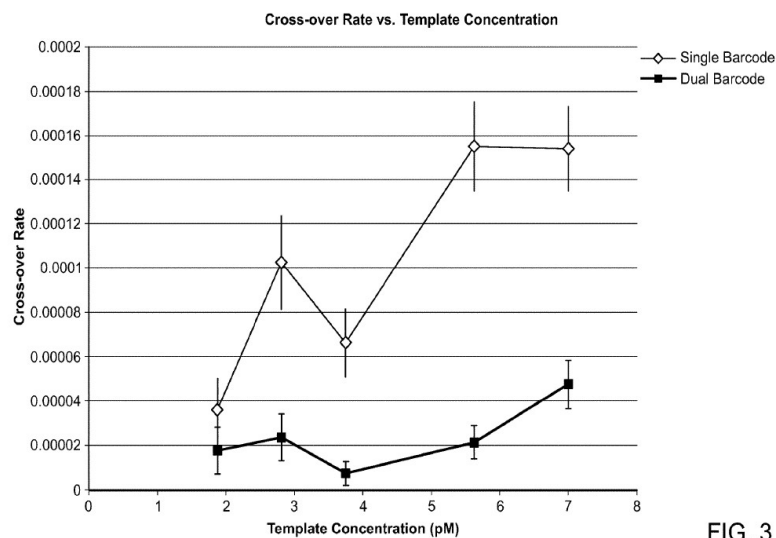


FIG. 3

Figure 16. Figure 3 of the '200 Patent
(EX1001 ('200 patent), Figure 3)

124. **Figure 16** compares the “crossover” rate between two sample preparation procedures: (i) a single barcode attached to template molecules before pooling and sequencing and (ii) a unique combination of two barcode sequences attached to template molecules before pooling and sequencing. EX1001 (’200 patent), 6:7-17, Figure 3. The ’200 patent states that **Figure 16** shows that “methods of the invention dramatically reduce false positives in a multiplex sequencing reaction.” EX1001 (’200 patent), 6:53-55.

125. But the ’200 patent provides no information about the nucleotide sequences of the barcodes used, specific barcode designs, or the considerations used in designing the barcodes—not in Figure 3, nor anywhere in the patent as a whole. *See generally* EX1001 (’200 patent). Likewise, the ’200 patent provides no information about the source of the samples or the library preparation protocols used in the study. *See generally* EX1001 (’200 patent). The patent also fails to mention which sequencing platform was used in the study. *See generally* EX1001 (’200 patent).

126. Notably, the ’200 patent provides no new guidance regarding laboratory chemistry for how to design, make, or use methods for dual-index sequencing. In fact, the ’200 patent states that “[n]umerous techniques are known for detecting sequence and for identifying barcodes and some are exemplified below. However, the exact means for detecting and compiling sequence data does not affect

the function of the invention.” EX1001 (’200 patent), 3:44-48. Further, the patent generally refers to existing scientific and patent publications in discussing: (1) the preparation of biological samples for sequencing (EX1001 (’200 patent), cols. 7-8); (2) barcode sequence design (EX1001 (’200 patent), cols. 8-10); (3) the amplification of barcoded templates (EX1001 (’200 patent), cols. 10-18); and (4) sequencing methods and platforms (EX1001 (’200 patent), cols. 18-20). *See also* EX1001 (’200 patent), 20:40-45 (“any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used”).

D. Prosecution History

127. From counsel, I understand that the ’200 patent belongs to the same patent family as U.S. Patent No. 9,163,281 (“the ’281 patent”; EX1030) and shares substantially the same specification with the ’281 patent. As illustrated below, the claimed subject matter of the ’200 patent is substantially similar to that of the ’281 patent, except that the ’200 patent requires the two identifiers to be “different” whereas the ’281 patent requires the two identifiers to be “the same.”

’281 Patent, Claims 1 & 2	’200 Patent, Claim 1
1. A method for validating the integrity of a nucleic acid template in a multiplex sequencing reaction, the method comprising: obtaining a template nucleic acid;	1. A method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction, the method comprising:

'281 Patent, Claims 1 & 2	'200 Patent, Claim 1
incorporating at least two of the same sequence identifiers into the template, wherein the sequence identifiers comprise four or more nucleotides; sequencing the template to generate sequence reads, and analyzing the sequence reads to determine if the two incorporated sequence identifiers remain associated with the template, thereby validating the integrity of the nucleic acid template in a multiplex sequencing reaction.	i) detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest, wherein at least a first identifier sequence is incorporated into a 5' portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3' portion of the nucleic acid analyte of interest, and wherein the first and second identifier sequences have four or more nucleotides and <i>are different</i>; ii) sequencing the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence; and iii) validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from

'281 Patent, Claims 1 & 2	'200 Patent, Claim 1
2. The method according to claim 1, wherein a first identifier is incorporated into a 5' portion of the template and a second identifier is incorporated into a 3' portion of the template.	sequence analysis of the multiplex sequencing reaction.

1. Prosecution History of the '281 Patent

128. The '281 patent was filed with claims directed to a “method for validating the integrity of a nucleic acid template in a multiplex sequencing reaction” by “incorporating at least two sequence identifiers into the template.” EX1026 ('281 patent file history), 34. The claims were later amended to include an additional step of “analyzing the sequence reads to determine if the two incorporated sequence identifiers remain associated with the template, thereby validating the integrity of the nucleic acid template in a multiplex sequencing reaction.” EX1026 ('281 patent file history), 146-148, 165-176.

129. The Examiner rejected the amended claims, contending that *Parameswaran* anticipates the claims. EX1026 ('281 patent file history), 184-185. In response, the applicant amended the claims to “incorporate[] at least two of the *same* sequence identifiers into the template.” EX1026 ('281 patent file history), 262-274. The applicant also argued that the amended claims are different from

Parameswaran because *Parameswaran* “requires attaching barcodes that are not the same.” EX1026 (’281 patent file history), 394-95 (emphasis in original).

130. The Examiner maintained the anticipation rejection, asserting that at least one pair of 10-nucleotide barcode sequences used in *Parameswaran* shared three identical nucleotides. EX1026 (’281 patent file history), 415-417. In response, the applicant amended the claims by adding a length limitation, i.e., that the claimed two sequence identifiers must be “the same” and “comprise four or more nucleotides.” EX1026 (’281 patent file history), 444-456. The applicant also argued that *Parameswaran* does not anticipate the amended claims because it “reports a method of sequencing that involves attaching a forward barcode and a reverse barcode to a nucleic acid template, the forward and reverse barcodes having *different designs* from each other.” EX1026 (’281 patent file history), 452-453.

131. The Examiner allowed the claims and noted that the “amendments overcame the previously presented rejections.” EX1026 (’281 patent file history), 458-464.

2. Prosecution History of the ’200 Patent

132. The ’200 patent issued without facing any anticipation or obviousness rejection over *Parameswaran* or any other prior art reference. In allowing the ’200 patent, the Examiner noted that “no references were found teaching or suggesting” the claims of the ’200 patent. EX1007 (’200 patent file

history), 395-401. The Examiner apparently overlooked her prior observation that *Parameswaran* discloses the use of two identifier sequences that are “different” to validate the integrity of a nucleic acid template in a multiplex sequencing reaction.

E. Claim Construction

133. In my analysis, I have read all claim terms as having their ordinary meaning as understood by a POSA.

VIII. THE CHALLENGED CLAIMS ARE UNPATENTABLE OVER THE PRIOR ART

134. In my opinion, claims 1-6 are anticipated by *Gloor* and *Parameswaran*. Further, claims 1-6 would have been obvious to a POSA at the earliest “effective filing date” of the ’200 patent.

A. Ground 1: Claims 1-6 Are Anticipated by *Gloor*

135. As explained below, claims 1-6 of the ’200 patent are anticipated by *Gloor* (EX1005) because *Gloor* discloses each and every limitation of claims 1-6. *Supra* § VI.B.

1. Claim 1 is Anticipated by *Gloor*

- a) 1[pre]: “A method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction”

136. *Gloor* discloses this limitation. Using Illumina sequencing, *Gloor* sequenced nucleic acids derived from 272 clinical samples that had been amplified using “combinatorial sequence tags attached to PCR primers.” EX1005

(Gloor), Abstract. Each nucleic acid analyte of interest was attached a pair of sequence tags (i.e. the “identifier sequences” used in claim 1[a]). EX1005 (Gloor), 7 (Figure 2). These combinatorial pairs of sequence tags were unique, i.e., nucleic acid analytes of interest from each sample had their own pairs of identifier sequences, and no two samples shared the same pair. EX1005 (Gloor), Table 2 (listing sequence tags used to create combinatorial pairs); *see also* Ex1005 (Gloor), 2 (“[w]e used paired-end sequencing in combination with unique sequence tags”).

137. *Gloor* used these unique pairs of sequence tags to validate the sequences of the nucleic acid analytes of interest. In *Gloor*, the paired-end reads for a single nucleic acid analyte of interest included an overlap between the sequence reads. In other words, two overlapping sequence reads were combined to provide the sequence of a single nucleic acid analyte of interest. EX1005 (Gloor), 7. *Gloor* discloses that, of the 18,047,860 reads obtained, 14,960,488 reads had a perfect overlap of between 10-50 nucleotides. EX1005 (Gloor), 7.

138. From these sequences with perfect overlap, *Gloor* detected the presence of both identifier sequences, and extracted sequences “with proper primer sequences and correct sequence tags.” EX1005 (Gloor), 7. For “the remainder of the analysis,” *Gloor* only used sequences with “two valid sequence tags and both primer sequences.” EX1005 (Gloor), 7-8. *Gloor* concludes that its study showed that “sequence clustering of the large number of reads derived from Illumina sequencing

would be more accurate” by accounting for, e.g., the underlying error rates. EX1005 (Gloor), 22-23. In other words, *Gloor* recognizes that its dual-indexing method led to “more accurate” assignments of sequences in Illumina multiplex sequencing.

139. *Gloor* also examined the sequences with “incorrect sequence tags” that were excluded from further analysis. EX1005 (Gloor), 7-8. *Gloor* reports that many of those sequences completely “lack[ed] the left or right end sequence tag.” EX1005 (Gloor), 7-8. In some cases, they contained sequence tags that were truncated, included additional bases, or had misincorporated bases. EX1005 (Gloor), 7-8. *Gloor* explains that by imposing “the strict requirement for proper sequence tag and primer sequences on the left and right ends,” many erroneous sequences were eliminated. EX1005 (Gloor), 22.

140. In sum, *Gloor* used Illumina sequencing to sequence nucleic acid analytes of interest from multiple samples simultaneously (i.e., multiplex sequencing), validated the sequences of those analytes by detecting the presence of two identifier sequences in each sequence, and excluded sequences that contained only one tag or incorrect tags. Thus, *Gloor* discloses a method of “validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction.”

b) 1[a]: “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest”

141. *Gloor* discloses this limitation. First, *Gloor* discloses “two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest.” As discussed above (*supra* § VI.B), *Gloor* extracted nucleic acids from 272 clinical samples and amplified them using “combinatorial sequence tags attached to PCR primers.” EX1005 (*Gloor*), Abstract. *Gloor* states that the tags, referred to as the “L-tag” and “R-tag,” were added to the forward and reverse PCR primers before Illumina paired-end sequencing, respectively. EX1005 (*Gloor*), 2, 6. While multiple samples used the same left or right identifier sequence, the *pair* of identifier sequences was unique to that sample (i.e., non-unique dual-indexing). EX1005 (*Gloor*), Abstract (explaining that 272 clinical samples were attached to “combinatorial sequence tags”), Table 2 (listing 18 left identifier sequences and 16 right identifier sequences), 2 (“[w]e used paired-end sequencing in combination with unique sequence tags”). Thus, *Gloor* discloses “two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest.”

142. *Gloor* also discloses a method in which the presence of those identifier sequences is detected. *Gloor* sequenced the amplified nucleic acid analytes and detected the presence of both identifier sequences on the 5’ and 3’ ends of each nucleic acid analyte of interest. EX1005 (*Gloor*), 7 (identifying “sequences

contain[ing] two valid sequence tags and both primer sequences”); *see also* EX1005 (Gloor), 6 (Figure 2) (illustrating the detection of sequences, each containing two identifier sequences).

143. As such, *Gloor* discloses a method of detecting the presence of at least two identifier sequences, where the pair of identifiers are “uniquely associated with the nucleic acid analyte of interest.”

- c) **1[b]: “at least a first identifier sequence is incorporated into a 5’ portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3’ portion of the nucleic acid analyte of interest”**

144. *Gloor* discloses this claim limitation. Figure 2 of *Gloor* depicts the arrangement of the two identifier sequences and the nucleic acid analyte of interest. EX1005 (Gloor), 7 (Figure 2). As annotated in **Figure 17** below, *Gloor* discloses that the “L-tag” (corresponding to the “first identifier sequence” in claim 1[b]) was incorporated into the 5’ portion of the nucleic acid analyte of interest, and the “R-tag” (corresponding to the “second identifier sequence” in claim 1[b]) was incorporated into the 3’ portion of the nucleic acid analyte of interest. EX1005 (Gloor), 2 (“combination of sequence tags at each end of the overlapped reads”), 5-6 (describing “left-side” and “right-side” sequence tags incorporated into the “end of the PCR primers”).

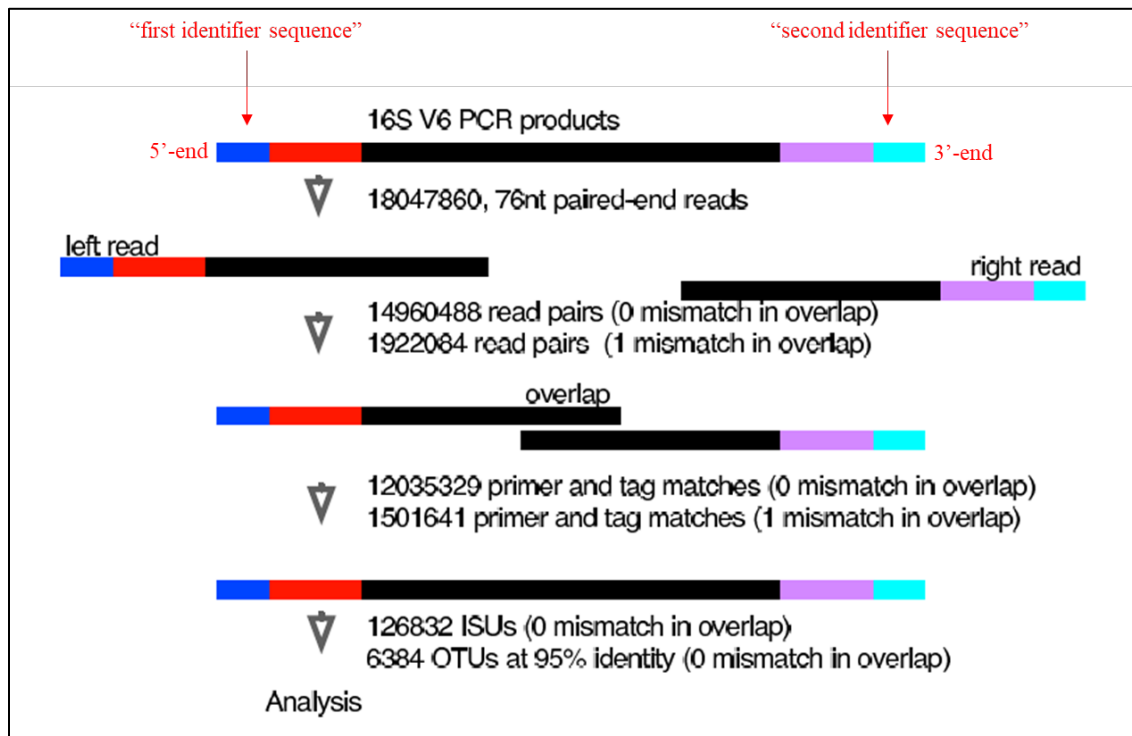


Figure 17. Paired-End Sequencing
(EX1005 (Gloor), 7 (Figure 2) (annotations added in red))

As such, *Gloor* discloses that “the first identifier sequence is incorporated into a 5’ portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3’ portion of the nucleic acid analyte of interest.”

d) 1[c]: “the first and second identifier sequences have four or more nucleotides and are different”

145. *Gloor* discloses this claim limitation. The sequence tags used in *Gloor* were “between 3 and 6 nucleotides long.” EX1005 (Gloor), 5. As shown in Table 2 and reproduced as **Figure 18** below, the majority of these sequence tags had four or more nucleotides. EX1005 (Gloor), 6 (Table 2).

L-tag	Name	R-tag	Name
catgcg	0-v6L	cgcatg	0-v6R
gcagt	1-v6L	actgc	1-v6R
tagct	2-v6L	agcta	2-v6R
gactgt	3-v6L	acagtc	3-v6R
cgtcga	4-v6L	tcgacg	4-v6R
gtcgc	5-v6L	gcgac	5-v6R
acgta	6-v6L	tacgt	6-v6R
cactac	7-v6L	gtagtg	7-v6R
tgac	8-v6L	gtca	8-v6R
agta	9-v6L	tact	9-v6R
atga	10-v6L	tcat	10-v6R
tgca	11-v6L	tgca	11-v6R
act	12-v6L	agt	12-v6R
tcg	13-v6L	cga	13-v6R
gta	14-v6L	tac	14-v6R
cta	15-v6L	tag	15-v6R
tga	16-v6L		
gcta	17-v6L		

Figure 18. Identifier Sequences Used in *Gloor*
(EX1005 (Gloor), 6 (Table 2))

146. As shown above, the two identifier sequences used in many pairs were different from each other. For example, the “0-v6L” left tag has six nucleotides (“catgcg”) and the “1-v6R” right tag has five nucleotides (“actgc”) and these tags are different from each other. Likewise, the “0-v6L” left tag has six nucleotides (“catgcg”) and the “2-v6R” right tag has five nucleotides (“agcta”) and these tags are different from each other.

147. As such, *Gloor* discloses many identifier sequences constituting “first and second identifier sequences [that] have four or more nucleotides and are different.”

e) **1[d]: “sequencing the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence”**

148. *Gloor* discloses this claim limitation. *Gloor* states that “[a]mplified PCR products are sequenced using an Illumina paired-end protocol to generate millions of overlapping reads.” EX1005 (*Gloor*), Abstract. As illustrated in **Figure 17** (*supra* § VIII.A.1.c)), each sequence read contained the sequence of the left tag and the nucleic acid analyte of interest, or the right tag and the nucleic acid analyte of interest. EX1005 (*Gloor*), 6 (Figure 2). *Gloor* compiled the sequence of each analyte of interest based on substantially overlapping sequence reads. EX1005 (*Gloor*), 7. Thus, the sequences of the analytes each contained the left tag (corresponding to the “first identifier sequence” in claim 1[d]), the nucleic acid analyte of interest, and the right tag (corresponding to the “the second identifier sequence” in claim 1[d]). EX1005 (*Gloor*), 6 (Figure 2).

149. As such, *Gloor* discloses a method that involves “sequencing the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence.”

- f) **1[e]: “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction”**

150. *Gloor* discloses this limitation. As discussed above (*supra* §§ VIII.A.1.a), VIII.A.1.c), VIII.A.1.e)), *Gloor* disclosed a method of sequencing a nucleic acid analyte of interest wherein each analyte was flanked by a uniquely associated pair of left and right tags. EX1005 (*Gloor*), 6, 22. *Gloor* used an “Illumina paired-end protocol to generate millions of overlapping reads.” EX1005 (*Gloor*), Abstract. *Gloor* compiled those sequence reads based on substantial overlap and “extract[ed] overlapping reads with proper primer sequences and correct sequence tags.” EX1005 (*Gloor*), 7. Of the about 18 million sequence reads collected, *Gloor* identified about 12 million sequences that “contained two valid sequence tags and both primer sequences,” which “form[ed] the basis of the remainder of the analysis.” EX1005 (*Gloor*), 7. *Gloor* concludes that its dual-indexing method led to “more accurate” sequence assignments. EX1005 (*Gloor*), 22-23. As such, *Gloor* discloses “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers.”

151. *Gloor* also examined sequences with “incorrect sequence tags” that were excluded from the remainder of the analysis, and reports that these sequences

completely “lack[ed] the left or right end sequence tag” (i.e., sequences having only one identifier sequence) or contained truncated tags, tags with additional bases, or tags with misincorporated bases (i.e., sequences with incorrect pair of identifiers). EX1005 (Gloor), 7-8; *see also* EX1005 (Gloor), 22 (erroneous sequences were removed from further analysis by applying “the strict requirement for proper sequence tag and primer sequences on the left and right ends”). As such, *Gloor* discloses “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction.”

152. *Gloor* therefore anticipates claim 1 because it discloses each and every element of the claim.

2. Dependent Claims 2-6 Are Anticipated by *Gloor*

153. For the reasons stated above in Section VIII.A.1 and below, dependent claims 2-6 are also anticipated by *Gloor*.

a) Dependent Claims 2 & 3

154. Dependent claim 2 recites the method of claim 1, “wherein the identifier sequences are nucleic acid sequence tags.” *Gloor* anticipates dependent claim 2. Dependent claim 3 recites the method of claim 2, “wherein the identifiers

are barcode sequences.” The ’200 patent states that nucleic acid barcodes can be used as identifier sequences. EX1001 (’200 patent), 2:16-20.

155. *Gloor* discloses a barcoding approach using a “combination of sequence tags at each end” of PCR primers. EX1005 (*Gloor*), 2. *Gloor* also discloses that the sequence tags were nucleic acid sequence tags. EX1005 (*Gloor*), 5-6 (stating that its identifier sequences were “between 3 and 6 nucleotides long”). As shown in Table 2, the identifier sequences were nucleic acid sequence tags. EX1005 (*Gloor*), 5-6 (Table 2). Thus, *Gloor* discloses identifier sequences that are nucleic acid sequence tags and barcode sequences. As such, *Gloor* discloses dependent claims 2 and 3.

b) Dependent Claim 4

156. Dependent claim 4 recites the method of claim 3, “wherein the barcode sequences are from 4 to 20 nucleotides.” *Gloor* anticipates dependent claim 4.

157. *Gloor* provides the sequences of left- and right-side identifier sequences in Table 2. EX1005 (*Gloor*), 6 (Table 2). As shown in Table 2, the majority of the barcode sequences *Gloor* used had four or more nucleotides. As noted above, I understand that a reference disclosing use of one composition within the scope of the claimed subject matter anticipates the claim. As such, *Gloor* discloses dependent claim 4.

c) Dependent Claim 5

158. Dependent claim 5 recites the method of 4, “wherein t [sic] barcode sequences are designed to have no homopolymer regions.” *Gloor* anticipates dependent claim 5.

159. *Gloor* provides the sequences of left- and right-side identifier sequences in Table 2. EX1005 (*Gloor*), 6 (Table 2). As shown in Table 2, *Gloor* utilized barcode sequences without homopolymeric runs. EX1005 (*Gloor*), 6 (Table 2). For instance, the sequence “catgcg” corresponding to “0-v6L” and the sequence “actgc” corresponding to “1-v6R” contain no consecutively repeated nucleotides. EX1005 (*Gloor*), 6 (Table 2) (showing that the identifier sequences contain no consecutively repeated nucleotides). Thus, the barcode sequences were designed to have no homopolymer regions. As such, *Gloor* discloses dependent claim 5.

d) Dependent Claim 6

160. Dependent claim 6 recites the method of claim 1, “wherein the nucleic acid analyte of interest is on the surface of a flow cell.” *Gloor* anticipates dependent claim 6.

161. *Gloor* discloses use of Illumina sequencing, which involved attaching nucleic acid analytes on the surface of a flow cell. *Supra* §§ VI.B, V.B.3; EX1025 (Illumina Technology Spotlight), 1-2. As such, *Gloor* discloses dependent claim 6.

B. Ground 2: Claims 1-6 Are Anticipated by *Parameswaran*

162. As explained below, claims 1-6 of the '200 patent are anticipated by *Parameswaran* (EX1004) because *Parameswaran* discloses each and every limitation of claims 1-6. *Supra* § VI.A.

1. Claim 1 Is Anticipated by *Parameswaran*

- a) 1[pre]: **“A method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction”**

163. *Parameswaran* discloses this limitation. Using a Roche 454 platform, *Parameswaran* sequenced “DNA from two independent sets of 25 barcoded libraries.” EX1004 (*Parameswaran*), 7. “[E]ach sample [was] associated with two uniquely designed, 10-nucleotide barcodes.” EX1004 (*Parameswaran*), 2. *Parameswaran* further “require[d] a match to the 10-base barcodes and additional three bases of the fixed linker” to facilitate “unambiguous assignment of nucleic acid sequences from a mixture of libraries from up to 48 different samples.” EX1004 (*Parameswaran*), 8. *Parameswaran* describes the combination of the 10-nucleotide barcode plus three nucleotides from the linker sequence as a 13-nucleotide “barcode motif” (i.e., the “identifier sequence” used in claim 1[a]). EX1004 (*Parameswaran*), 6.

164. *Parameswaran* reports that around 94% of sequences retained a correct pair of both forward and reverse barcode motifs. EX1004 (*Parameswaran*), 7. That

is, the forward and reverse barcode motifs matched the sequences of the 10-nucleotide barcodes and three nucleotides from the adjacent linkers and, additionally, the forward barcode motif was paired with the correct reverse barcode motif. *Parameswaran* reports that sequences with two matching barcode motifs were accurately and unambiguously assigned to their originating sample, i.e., their sequences were validated. EX1004 (*Parameswaran*), 8 (“any sequence with both 5’ and 3’ barcode motifs will be placed with reasonable accuracy into the correct bin”).

165. *Parameswaran* also reports that about 5-6% of sequences had a perfect barcode motif at only one end. EX1004 (*Parameswaran*), 7. Those sequences included sequences with a partial or no identifier barcode motif on the other end (i.e., sequences with only one identifier) or sequences with a pair of mismatched (or “misassigned”) barcode motifs. EX1004 (*Parameswaran*), 7. *Parameswaran* states that the latter subset of sequences has “perfect identity to an unexpected motif, resulting in an unassigned 5’-3’ motif combination (perfect, unexpected occurrence).” EX1004 (*Parameswaran*), 7. In those sequences, both the forward and reverse barcode motifs matched the sequences of valid barcode motifs used in the multiplex reaction, but the forward barcode motif was paired with an incorrect reverse barcode motif. In other words, both the forward and reverse barcode motifs corresponded to valid barcode motifs, but each was associated with a different

sample.³ As *Parameswaran* explains, various errors could be responsible for sequences with one barcode motif or mismatched barcode motifs, including, for example, cross-contamination of DNA templates, errors in sequencing-by-synthesis, errors in emulsion PCR, and barcoded primer synthesis errors. EX1004 (*Parameswaran*), 7.

166. *Parameswaran* refers to the sequences that contain “unexpected perfect pairings of unmatched forward and reverse barcodes” as “false-discover[ies],” representing misassignments of sequences to incorrect samples. EX1004 (*Parameswaran*), Abstract, 8. *Parameswaran* states that a “remarkably low misassignment rate (0.00042% and 0.0044% for two independent experiments) was obtained by requiring a match to the 10-base barcodes and additional three bases of the fixed linker” (i.e., the forward and reverse barcode motifs). EX1004 (*Parameswaran*), 8. *Parameswaran* further states that “[t]he low false-discovery rate

³ Referring to the illustrative example that I provided above (*supra*, § V.B.6), if the correct barcode pair for a given template sequence was “H-8,” a “misassigned” barcode motif in *Parameswaran* would have the 5’ identifier “H” paired with a different 3’ identifier (e.g., H-2 or H-14) or would have the 3’ identifier “8” paired with a different 5’ identifier (e.g., A-8 or J-8).

for perfect matches strengthens the argument that any sequence with both 5' and 3' barcode motifs will be placed with reasonable accuracy into the correct bin.” EX1004 (Parameswaran), 8. In other words, *Parameswaran* teaches that lowering the false discovery (misassignment) rate allows sequences to be assigned to correct samples more accurately.

167. In sum, *Parameswaran* sequenced multiple nucleic acid analytes of interest simultaneously (i.e., multiplex) using Roche 454 sequencing and validated the sequences of those analytes by detecting the presence of two identifier sequences in each sequence. *Parameswaran* separately identified erroneous sequences with only one barcode motif or a pair of mismatched barcode motifs, the latter of which was recognized as a “misassignment.” EX1004 (Parameswaran), 2, 8. *Parameswaran* therefore discloses a “method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction.”

b) 1[a]: “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest”

168. *Parameswaran* discloses this limitation. First, *Parameswaran* discloses “two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest.” As discussed above (*supra* § VI.A), *Parameswaran* attached “two uniquely designed, 10-nucleotide barcodes” to nucleic acid analytes of interest using nucleotide linkers. EX1004 (Parameswaran), 2. *Parameswaran* “require[d] a

match to the 10-base barcodes and additional three bases of the fixed linker” (“barcode motifs”) to facilitate “unambiguous assignment of nucleic acid sequences from a mixture of libraries from up to 48 different samples.” EX1004 (Parameswaran), 8. As shown in Supplementary Table 1 of *Parameswaran*, no barcode motif repeated across samples, i.e., each pair of barcode motifs was “uniquely” associated with nucleic acid analytes of interest from a specific sample. EX1004 (Parameswaran), 14-17 (Supplementary Table 1).

169. *Parameswaran* also discloses a method in which the presence of two or more identifier sequences was detected. *Parameswaran* amplified the nucleic acid analytes of interest containing a pair of barcode motifs and sequenced them. EX1004 (Parameswaran), 7-8. *Parameswaran* detected about 94% of sequences having a pair of perfect forward and reverse barcode motifs associated with its originating sample. EX1004 (Parameswaran), 7.

170. As such, *Parameswaran* discloses “detecting the presence of at least two identifier sequences, each of which is uniquely associated with the nucleic acid analyte of interest.”

- c) 1[b]: “at least a first identifier sequence is incorporated into a 5’ portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3’ portion of the nucleic acid analyte of interest”

171. *Parameswaran* discloses this claim limitation. *Parameswaran* states that each nucleic acid analyte was attached to a pair of 13-nucleotide “barcode motifs,” each of which was made of a 10-nucleotide barcode and three nucleotides from an adjacent linker. EX1004 (*Parameswaran*), 2, 7. **Figure 19** provides a schematic of the template nucleic acids with barcodes (i.e., F-Barcode, R-Barcode) and linkers (collectively, “barcode motifs” or “identifier sequences”) incorporated into the 5’ and 3’ ends of each nucleic acid analyte of interest.

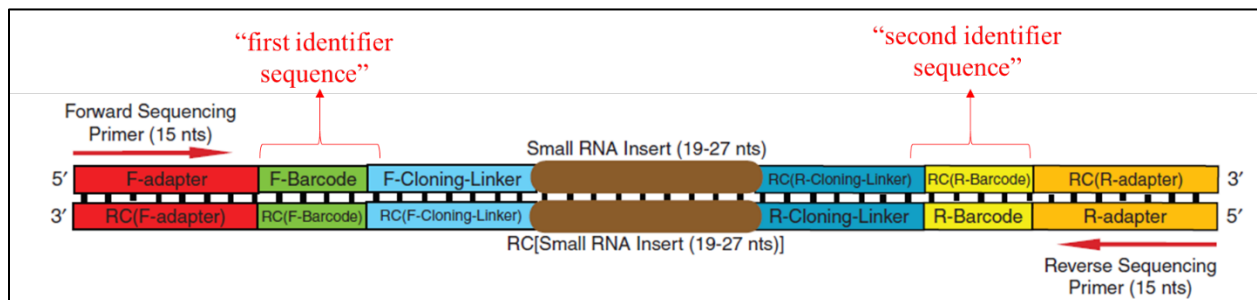


Figure 19. Design of Forward and Reverse Primers
(EX1004 (*Parameswaran*), 3 (Figure 1C) (annotations added))

The forward barcode motif, corresponding to the “first identifier sequence” in claim 1[c], consisted of the “F-Barcode” and three nucleotides from the adjacent linkers to form barcode motifs at the 5’ end of the nucleic acid analyte. EX1004 (*Parameswaran*), 5. The reverse barcode motif, corresponding to the “second identifier sequence” in claim 1[c], consisted of the “R-Barcode” and three

nucleotides from the adjacent linkers to form barcode motifs at the 3' end of the nucleic acid analyte. EX1004 (Parameswaran), 5.

172. As such, *Parameswaran* discloses a method that involves “at least a first identifier sequence is incorporated into a 5' portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3' portion of the nucleic acid analyte of interest.”

d) 1[c]: “the first and second identifier sequences have four or more nucleotides and are different”

173. *Parameswaran* discloses this limitation. *Parameswaran* explains that each nucleic acid analyte of interest was attached to a pair of “10-nucleotide” barcodes, which “differ[ed] from each other at multiple positions (by at least 4 nt).” EX1004 (Parameswaran), 2. To ensure that all samples were uniquely dual-indexed, *Parameswaran* combined three nucleotides from the adjacent linkers with the forward and reverse barcodes, creating a pair of 13-nucleotide “barcode motifs.” EX1004 (Parameswaran), 7; *see also* EX1004 (Parameswaran), 14-17 (Supplementary Table 1) (showing that each barcode motif was different from one another, and no barcode motif was repeated across samples).

174. Thus, the forward and reverse barcode motifs used in *Parameswaran* are the “first and second identifier sequences [that] have four or more nucleotides and are different.”

- e) 1[d]: “sequencing the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence”

175. *Parameswaran* discloses this limitation. *Parameswaran* describes a multiplex sequencing reaction where nucleic acid analytes of interests containing the first and second identifier sequences were sequenced. Figure 1C in *Parameswaran* depicts sequences generated from this reaction. As reproduced again in **Figure 19** below, each sequence included the forward barcode motif (corresponding to the “first identifier sequence” in claim 1[d]), the nucleic acid analyte of interest, and the reverse barcode motif (corresponding to the “second identifier sequence” in claim 1[d]). EX1004 (*Parameswaran*), 3 (Figure 1C).

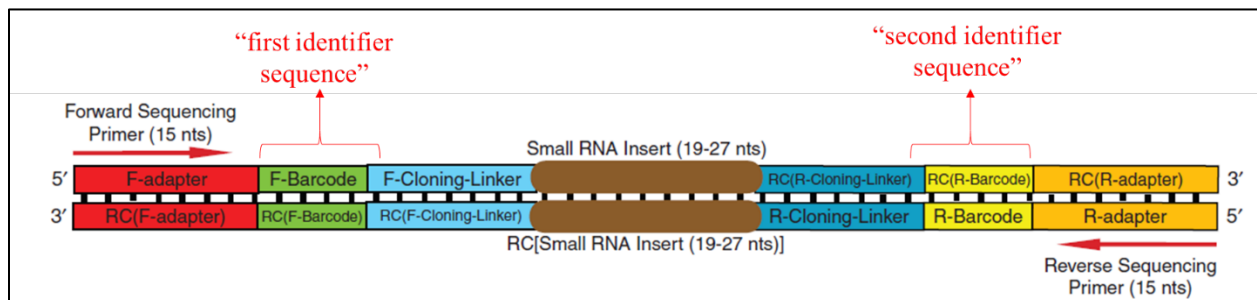


Figure 19. Design of Forward and Reverse Primers
(EX1004 (*Parameswaran*), 3 (Figure 1C) (annotations added))

As such, *Parameswaran* discloses “sequencing the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence.”

- f) **1[e]: “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction”**

176. *Parameswaran* discloses this limitation. As discussed above (*supra* §§ VIII.B.1.a), VIII.B.1.c) VIII.B.1.e)), *Parameswaran* describes a method for sequencing nucleic acid analytes from multiple samples that are each associated with a unique pair of forward and reverse barcode motifs, analyzing the barcode motifs present in each sequence, and validating the sequence of the nucleic acid analyte of interest. EX1004 (*Parameswaran*), 7 (reporting that about 94% of sequences had perfect forward and reverse barcode motifs on both ends); *see also* EX1004 (*Parameswaran*), 8 (“any sequence with both 5’ and 3’ barcode motifs will be placed with reasonable accuracy into the correct bin”). Thus, *Parameswaran* discloses “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers.”

177. While the majority of sequences reported in *Parameswaran* retained a correct pair of forward and reverse barcode motifs, about 5-6% of sequences contained only one perfect barcode motif, or a pair of mismatched (or “misassigned”) barcode motifs. EX1004 (*Parameswaran*), 7. *Parameswaran* refers to the sequences with misassigned barcode motif pairs as “false-discover[ies],”

constituting “misassignment[s]” of sequences to incorrect samples. EX1004 (Parameswaran), Abstract, 6, 8. *Parameswaran* explained that these sequences with one barcode motif or mismatched barcode motifs may have arisen from various errors such as cross-contamination of DNA templates and errors during sequencing-by-synthesis and emulsion PCR. EX1004 (Parameswaran), 7.

178. *Parameswaran* states that “[f]or our purposes, all sequences with a match to one of the two expected sequence motifs were considered as true members of the corresponding dataset.” EX1004 (Parameswaran), 8. But *Parameswaran* also explains that sequences with mismatched barcode motifs could be attributed to various errors and refers to such sequences as “false-discovery,” representing misassignments of sequences to incorrect samples. EX1004 (Parameswaran), 7, Abstract, 8. *Parameswaran* reports that the “false-discovery rate” in two independent experiments was 0.00042% and 0.0044%, reflecting a “remarkably low misassignment rate.” EX1004 (Parameswaran), 7-8.

179. *Parameswaran* thus recognizes that mismatched sequences are erroneous and that reducing the rate of such “false discovery” would decrease the rate at which sequences are assigned to incorrect samples. EX1004 (Parameswaran), 7-8. *Parameswaran* calls out the low incidence of such errors when requiring a match to the 13-base barcode motif as an advantage of its unique dual-barcoding method. EX1004 (Parameswaran), 8. Further, *Parameswaran* links

sequencing accuracy with the presence of both barcode motifs, stating that “any sequence with both 5’ and 3’ barcode motifs will be placed with reasonable accuracy into the correct bin.” EX1004 (Parameswaran), 8. Thus, *Parameswaran* discloses a method where erroneous sequences containing one identifier or an incorrect pair of identifiers can be discarded as erroneous, resulting in decreased misassignment of sequences to the corresponding samples.

180. As such, *Parameswaran* discloses “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction.”

181. As such, I conclude *Parameswaran* anticipates claim 1 as it discloses each and every element of the claim.

2. Dependent Claims 2-6 Are Anticipated by *Parameswaran*

182. For the reasons stated above in Section VIII.A.1 and below, dependent claims 2-6 are anticipated by *Parameswaran*.

a) Dependent Claims 2 & 3

183. Dependent claim 2 recites the method of claim 1, “wherein the identifier sequences are nucleic acid sequence tags.” *Parameswaran* anticipates dependent claim 2. Dependent claim 3 recites the method of claim 2, “wherein the

identifier sequences are barcode sequences.” The ’200 patent states that nucleic acid barcodes can be used as identifier sequences. EX1001 (’200 patent), 2:16-20.

184. *Parameswaran* describes a “high-information-content barcoding approach in which each sample is associated with two uniquely designed, 10-nucleotide barcodes.” EX1004 (*Parameswaran*), 2 (emphasis added). *Parameswaran* explains that the ten nucleotide barcodes (made of A, G, C, and T) and three nucleotides from adjacent linkers (also made of A, G, C, and T) were used as “barcode motifs,” or identifier sequences. EX1004 (*Parameswaran*), 5 (Table 2), 7, 14-17 (Suppl. Table 1). The barcode motifs in *Parameswaran* thus are nucleic acid sequence tags and barcode sequences. As such, *Parameswaran* discloses dependent claims 2 and 3.

b) Dependent Claim 4

185. Dependent claim 4 recites the method of claim 3, “wherein the barcode sequences are from 4 to 20 nucleotides.” *Parameswaran* anticipates dependent claim 4.

186. *Parameswaran* explains its barcode motifs include “10-nucleotide barcodes” and three nucleotides from adjoining linkers (totaling 13 nucleotides in length). EX1004 (*Parameswaran*), 2, 7-8. As such, *Parameswaran* discloses dependent claim 4.

c) Dependent Claim 5

187. Dependent claim 5 recites the method of claim 4, “wherein t [sic] barcode sequences are designed to have no homopolymer regions.” *Parameswaran* anticipates dependent claim 5.

188. *Parameswaran* describes that its “barcode design [] facilitated unambiguous assignment, with care taken to avoid homopolymeric runs.” EX1004 (*Parameswaran*), 4. *Parameswaran* further explains “[t]his design avoids significant problems due to error rate at homopolymeric runs of nucleotides.” EX1004 (*Parameswaran*), 4. As such, *Parameswaran* discloses dependent claim 5.

d) Dependent Claim 6

189. Dependent claim 6 recites the method of claim 1, “wherein the nucleic acid analyte of interest is on the surface of a flow cell.” *Parameswaran* anticipates dependent claim 6.

190. Dependent claim 6 does not require the analyte of interest be attached to the surface of a flow cell. Instead, dependent claim 6 only requires the analyte to be “on the surface of a flow cell.” *Parameswaran* discloses a method of sequencing using a Roche 454 platform, which involves attaching a nucleic acid analyte of interest to beads and placing those beads in a flow cell. EX1010 (Kircher), 3-4; EX1011 (Metzker), 2-3; EX1012 (Shendure), 3-4; EX1028 (Roche 454 Manual), 6-8 (depicting bead-bound nucleic acid analytes placed in wells of a flow cell); *supra*

§§ V.B.2-3, VI.A. As such, *Parameswaran* discloses a method where “the nucleic acid analyte of interest is on the surface of a flow cell.”

C. Ground 3: Claims 1-6 Would Have Been Obvious Over *Parameswaran* and *Gloor*

191. As I explain above (*supra* §§ VIII.A-VIII.A.2.d), VIII.B-VIII.B.2.d)), *Parameswaran* and *Gloor* disclose each and every claim limitation of claims 1-6 of the '200 patent.

192. Further, the combination of *Parameswaran* and *Gloor* renders obvious claims of 1-6 of the '200 patent. *Supra* §§ VI.A-B. Both *Parameswaran* and *Gloor* would have motivated a POSA to validate the sequence of a nucleic acid analyte of interest by detecting the presence of two identifiers attached to each end of the analyte and excluding unwanted sequences derived from technical errors, including those that contain only one identifier or an incorrect pair of identifiers. Both *Parameswaran* and *Gloor* would have motivated a POSA to use Illumina sequencing, which was known to be more efficient, less expensive, and faster than Roche 454 sequencing. A POSA would have had a reasonable expectation of success in doing so because each reference empirically demonstrated the feasibility of using dual identifiers to exclude unwanted sequences during Roche 454 and Illumina multiplex sequencing. As such, claims 1-6 would have been obvious to a POSA as of December 23, 2010, over the combination of *Parameswaran* and *Gloor*.

1. Claim 1 Would Have Been Obvious Over *Parameswaran* and *Gloor*

- a) 1[pre]: “A method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction”**

193. As I explain above, both *Parameswaran* and *Gloor* disclose this limitation. *Supra* §§ VIII.B.1.a), VIII.A.1.a).

- b) 1[a]: “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest”**

194. As I explain above, both *Parameswaran* and *Gloor* disclose this limitation. *Supra* §§ VIII.B.1.b), VIII.A.1.b).

- c) 1[b]: “at least a first identifier sequence is incorporated into a 5’ portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3’ portion of the nucleic acid analyte of interest”**

195. As I explain above, both *Parameswaran* and *Gloor* disclose this limitation. *Supra* §§ VIII.B.1.c), VIII.A.1.c).

- d) 1[c]: “the first and second identifier sequences have four or more nucleotides and are different”**

196. As I explain above, both *Parameswaran* and *Gloor* disclose this limitation. *Supra* §§ VIII.B.1.d), VIII.A.1.d).

197. While *Gloor* included some identifier sequences that were three nucleotides in length (EX1005 (*Gloor*), 5 (Table 2)), it would have been obvious to

use sequence tags that were four or more nucleotides long. Most of the identifier sequences in *Gloor* were four or more nucleotides long. EX1005 (*Gloor*), 5 (Table 2). Further, *Parameswaran*, which used 13-nucleotide barcode motifs, explains that longer sequence identifiers allowed for “definitive assignment.” EX1004 (*Parameswaran*), 4, 7. Further, both *Parameswaran* and *Gloor* used “different” sequence identifiers, which could be beneficial for reducing sequencing error because the “same” sequence identifiers could cause analytes to be annealed together due to the perfect homology between the two sequence identifiers, creating unwanted chimeric molecules.

- e) **1[d]: “sequencing the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence”**

198. As I explain above, both *Parameswaran* and *Gloor* disclose this limitation. *Supra* §§ VIII.B.1.e), VIII.A.1.e).

- f) **1[e]: “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction”**

- i. ***Parameswaran* discloses claim 1[e]**

199. *Parameswaran* discloses this limitation. *Supra* § VIII.B.1.f). As discussed above, *Parameswaran* describes a method that determined that about 94% of sequences had perfect forward and reverse barcodes, thereby validating the

sequences for these nucleic acid analytes of interest. EX1004 (Parameswaran), 7, 8 (“any sequence with both 5’ and 3’ barcode motifs will be placed with reasonable accuracy into the correct bin”).

200. A POSA reading *Parameswaran* would have understood that it discloses sequencing methods configured to identify erroneous sequences that have only one perfect barcode motif, which can be kept or discarded depending on the goals of the project. *Parameswaran* detected sequences lacking any match to the expected barcode motifs, and “discarded” them from further analysis. EX1004 (Parameswaran), 8. While *Parameswaran* indicates that “[f]or our purposes, all sequences with a match to one of the two expected sequence motifs were considered as true members of the corresponding dataset,” *Parameswaran* also states that sequences with mismatched barcode motifs, which it calls a “false-discovery,” arise due to various errors and represent misassignments of sequences to incorrect samples. EX1004 (Parameswaran), 7. *Parameswaran* further discloses that it is desirable to reduce the “false-discovery rate,” noting that its 13-base barcode motifs led to a “remarkably low misassignment rate” of 0.00042% and 0.0044%. EX1004 (Parameswaran), 7-8. A POSA would have understood that *Parameswaran* identifies these sequences with mismatched barcode motifs as resulting from errors and leading to undesirable misassignment of sequences to incorrect samples. EX1004 (Parameswaran), 8. A POSA would have understood that such sequences

could be excluded to increase accuracy. Indeed, *Parameswaran* states that “any sequence with *both 5’ and 3’ barcode motifs* will be placed with reasonable accuracy into the correct bin.” EX1004 (*Parameswaran*), 8.

ii. *Gloor* discloses claim 1[e]

201. As explained above, *Gloor* discloses this limitation. *Supra* § VIII.A.1.f). Specifically, *Gloor* discloses the “extraction” of sequences with “correct sequence tags,” the identification of sequences that “contained two valid sequence tags,” and the exclusion of sequences with “incorrect sequence tags” from “the remainder of the analysis.” EX1005 (*Gloor*), 7. *Gloor* explains that the excluded sequences lacked the first or second sequence identifier (i.e., only contained one identifier) or contained errors in the first or second sequence identifiers (i.e., an incorrect pair of identifiers). EX1005 (*Gloor*), 7-8. *Gloor* thus discloses “validating the sequence of the nucleic acid analyte of interest” by: (1) “analyzing both identifiers” and (2) “excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction.”

2. A POSA Would Have Been Motivated to Combine *Parameswaran* and *Gloor*

202. A POSA in December 2010 would have been motivated to combine *Parameswaran* and *Gloor* for several reasons. *First*, a POSA reading *Parameswaran*

and *Gloor* would have understood that multiplex sequencing could result in erroneous sequences due to errors such as incorrect synthesis of barcoded primers, cross-contamination of template DNAs, molecular overlap of template nucleic acids, and other errors arising during PCR amplification. EX1004 (Parameswaran), 7; EX1005 (Gloor), 7-8. A POSA would have understood that erroneous sequences were undesirable, leading to “misassignment[s]” of sequences in multiplex sequencing. EX1004 (Parameswaran), Abstract; *see also* EX1005 (Gloor) 23 (disclosing that methods that accounted for errors would lead to “more accurate” Illumina sequencing). *Parameswaran* and *Gloor* both disclose the use of distinct pairs of sequence identifiers to reduce errors during multiplex sequencing. To reduce technical errors that occur during multiplex sequencing, a POSA would have been motivated to combine and use the dual indexing methods in *Parameswaran* and *Gloor*.

203. *Second*, a POSA would have been particularly motivated to use *Parameswaran*’s unique dual-indexing methods, which would allow the POSA to identify and exclude chimeric sequences. As of December 2010, a POSA would have known that chimeric sequences were problematic for next-generation sequencing platforms that used PCR amplification, including Illumina sequencing. *See supra* § V.B.6. *Parameswaran*’s unique dual-indexing methods identified various types of erroneous sequences, including those that contain a pair of mismatched barcode

motifs as would result from a chimeric molecule. EX1004 (*Parameswaran*), 7. A POSA would have known that the unique dual-indexing method disclosed in *Parameswaran* made it possible to compare pairs of identifiers assigned to nucleic acid samples to those actually detected in the sequences, allowing the POSA to identify and exclude chimeric sequences containing an incorrect pair of identifiers. In contrast, because *Gloor* repeated individual sequence identifiers across different samples, additional bioinformatic analyses were needed to identify and discard chimeric sequences. EX1005, 8-9 (identifying “putative chimeric sequences” by comparing obtained sequences themselves and against known reference sequences). A POSA would have understood that *Parameswaran*’s unique dual indexing methods identifies chimeric sequences and would have been motivated to use them to improve sequencing accuracy.

204. *Third*, a POSA would have been particularly motivated to apply the barcoding method of *Parameswaran* for Illumina sequencing as disclosed in *Gloor*, as Illumina sequencing was faster, more efficient, and less expensive. The throughput of Illumina sequencers was approximately 5,000 Mb/day, far higher than the 750 Mb/day throughput of Roche 454 sequencers. EX1010 (Kircher), 10 (Table 1). And the cost of Illumina sequencers was an order of magnitude lower than Roche 454 sequences—approximately \$0.50/Mb compared to \$20/Mb with Roche 454 sequencers. EX1010 (Kircher), 10 (Table 1). As *Gloor* states, the “Illumina

sequencing method generated ~ 12 million useable reads at a cost of ~ 0.03 cents per read, an approximate order of magnitude cheaper than the per-read cost of pyrosequencing.” EX1005 (Gloor), 2. These efficiency and cost advantages would have been strong motivators for anyone working in DNA sequencing to use Illumina sequencers. In my laboratory as of December 2010, we used Illumina sequencers over Roche 454 sequencers because, among other reasons, Illumina sequencers were faster and cheaper. EX1011 (Metzker), 34 (“Currently, the Illumina/Solexa Genome Analyzer (GA) dominates the [next-generation sequencing] market.”).

205. A POSA would have known that the greater efficiency and lower cost of Illumina sequencing are due in part to the fact that Roche 454 sequencing, as used in *Parameswaran*, required a multi-step process to amplify template nucleic acids, involving amplifying templates on beads, separating beads with the amplified templates, and placing selected beads into wells on a flow cell. EX1011 (Metzker), 3 (describing Roche 454 amplification workflow), EX1012 (Shendure), 1 (same). Further, this process was often performed manually, adding cost and increasing the time needed for sequencing. EX1028 (Roche 454 Manual), 6-9. In contrast, Illumina sequencing offered a simpler process in which amplifying templates on the surface of a flow cell was performed using an automated system, thereby decreasing cost and user time. EX1025 (Illumina Technology Spotlight), 1.

206. *Fourth*, a POSA would have been motivated to try Illumina sequencing in view of the express teachings in both *Parameswaran* and *Gloor*. As of December 2010, a POSA would have known that dual-indexing methods were platform-agnostic. While *Parameswaran* used a Roche 454 platform, *Parameswaran* explains that its barcodes “may also be compatible with other sequencing platforms,” and specifically identifies Illumina sequencing as a major high throughput sequencing platform. EX1004 (*Parameswaran*), 1, 8. Similarly, *Gloor* identifies “454 pyrosequencing” as a multiplex sequencing platform, but explains that Illumina sequencing had a “substantially lower error rate.” EX1005 (*Gloor*), 12. This express teaching in *Parameswaran* would have motivated a POSA to use Illumina sequencing, as already done in *Gloor*.

207. Further, given that there were only a limited number of platforms, it would have been obvious for a POSA to try *Parameswaran*’s unique dual-indexing method with one of the other sequencing platforms in place of Roche 454 sequencing, as *Parameswaran* suggests. In fact, researchers in the sequencing field had already done so. *See generally* EX1038 (*Maughan*). For example, *Maughan* sequenced two sets of nucleic acid libraries from 60 individual samples, using both Roche 454 and Illumina sequencers. EX1038 (*Maughan*), Abstract. *Maughan* used unique dual indexing for both platforms, where a unique “pair” of barcodes was incorporated into nucleic acids from each sample. EX1038 (*Maughan*), 4; *see also*

EX1038 (Maughan), 3 (Table 1). *Maughan* reinforces that it would have been obvious for a POSA to perform unique dual indexing using any known sequencer. This is consistent with the acknowledgement in the '200 patent that "essentially any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used." EX1001 ('200 patent), 20:40-45; *see supra* § VII.C.

208. As such, for at least these reasons, a POSA would have been motivated to combine *Parameswaran* and *Gloor* to validate the sequence of a nucleic acid analyte of interest by excluding undesirable sequence reads such as those containing only one identifier or an incorrect pair of identifiers.

3. A POSA Would Have Had a Reasonable Expectation of Success in Combining *Parameswaran* and *Gloor*

209. A POSA in December 2010 would have had a reasonable expectation of success in combining *Parameswaran* and *Gloor* to perform the claimed methods. As summarized above, both *Parameswaran* and *Gloor* disclosed detailed procedures for: (1) dual indexing nucleic acid analytes of interest from multiple samples; (2) sequencing them simultaneously using next-generation sequencers; (3) detecting first and second identifier sequences; and (4) validating the sequences of analytes of interest by excluding sequences containing only one identifier or incorrect pairs of identifiers. *Supra* §§ VIII.A.1, VIII.B.1, VIII.C.1. Following the teachings of

Parameswaran and *Gloor*, a POSA would have reasonably expected that the sequences of nucleic acid analytes of interest in a multiplex sequencing reaction would be successfully validated.

210. A POSA would have reasonably expected that the claimed methods could be successfully performed with either the Roche 454 or Illumina sequencing platforms. *Parameswaran* used unique dual indexing on a Roche 454 sequencer, explained that its methods would be compatible “with other sequencing platforms,” and identified Illumina sequencing as one of the “major high-throughput sequencing platforms.” EX1004 (*Parameswaran*), 1, 2, 8. *Gloor* similarly performed dual indexing, but did so on an Illumina sequencer. EX1005 (*Gloor*), Abstract, 6-8; *supra* §§ VI.B, VIII.A, VIII.A.2.d). In view of these successful demonstrations, a POSA would have had a reasonable expectation of success in combining *Parameswaran* and *Gloor*.

211. A POSA would also have had a reasonable expectation of success in applying *Parameswaran*’s unique dual indexing method on an Illumina sequencer as disclosed in *Gloor*. A POSA would have understood that performing the unique dual-indexing method described in *Parameswaran* with an Illumina sequencer as disclosed in *Gloor* simply required exchanging Roche-specific adapter sequences with Illumina-specific adapter sequences and other minor modifications. *Compare* EX1004 (*Parameswaran*), 14-15 (listing Roche-specific adapter sequences in the

same column as barcode and cloning linker sequences) *with* EX1006 (Bentley), 13-14 (listing sequences of “adapters for the preparation of paired read libraries” used for Illumina sequencing); *see* EX1015 (Meyer 2010), 1. A POSA would have appreciated that these modifications did not require changing the fundamental aspect of the dual barcoding methods disclosed in *Parameswaran* or *Gloor*.

212. *Maughan* empirically demonstrates this point. *Supra* § VIII.C.2. In *Maughan*, researchers performed unique dual-index sequencing using both Roche 454 and Illumina sequencers. EX1038 (Maughan), Abstract; EX1038 (Maughan), 4, 3 (Table 1). Notably, *Maughan* used the **same set** of barcode pairs for both platforms (EX1038 (Maughan), 3 (Table 1), illustrating that: (1) changing sequencers does not change the fundamental aspect of dual indexing and (2) the change can be successfully made with only minor modifications.

213. Notably, no novel or nonobvious chemistry would have been needed to perform *Parameswaran*’s unique dual-indexing on an Illumina sequencer. The ’200 patent acknowledges this fact and its disclosure refers to the general knowledge in describing the required (and non-novel) chemistry. *Supra* § VII.C. For example, it states that “[n]umerous techniques are known for detecting sequence and for identifying barcodes and some are exemplified below. However, the exact means for

detecting and compiling sequence data does not affect the function of the invention.”
EX1001 (’200 patent), 3:44-48.

214. A POSA would also have understood that the read lengths for Illumina sequencers disclosed in *Gloor* were long enough to perform the dual-indexing methods disclosed in *Parameswaran*. As of December 2010, Illumina sequencers were capable of “high-quality paired end reads of 150 base pairs (bp).” EX1027 (July 2010 Press Release), 1. The read lengths required for sequencing each *Parameswaran* template nucleic acid were between 71 and 81 nucleotides: two barcodes (10 nucleotides each), two cloning linkers (16-17 nucleotides each), and one small RNA insert (19-27 nucleotides). EX1004 (Parameswaran), 3 (Figure 1).⁴ *Gloor* successfully conducted Illumina paired-end sequencing on molecules using read lengths of 76 nucleotides, and moreover states that “Illumina sequencing is now capable [of] ~ 100 [nucleotide] long reads from each end of a DNA fragment.”

⁴ Adapter sequences are not included within the total read length because one adapter is annealed to the sequencing primer (and therefore not sequenced), and the other adapter is located sequentially after the two barcodes and small RNA insert and therefore not necessary for obtaining the sequence of the insert or assigning the sequencing reads to its original sample.

EX1005 (Gloor), 6-7, 2. As such, Illumina sequencers could sequence the read lengths of the molecules used in *Parameswaran*.

215. Based on the foregoing, I conclude that claim 1 of the '200 patent would have been obvious over the combination of *Parameswaran* and *Gloor*.

4. Dependent Claims 2-6 Would Have Been Obvious Over the Combination of *Parameswaran* and *Gloor*

216. For the reasons stated above in Section VIII.C.1 and below, the additional limitations of dependent claims 2-6 would have been obvious over the combination of *Parameswaran* and *Gloor*.

a) Dependent Claims 2 & 3

217. Dependent claim 2 recites the “method of claim 1, wherein the identifier sequences are nucleic acid sequence tags.” Dependent claim 3 recites the “method of claim 2, wherein the identifier sequences are barcode sequences.” As detailed above, *Parameswaran* and *Gloor* both disclosed methods of using identifier sequences that are nucleic acid sequence tags and barcode sequences. *Supra* §§ VIII.B.2.a) (*Parameswaran*), VIII.A.2.a) (*Gloor*). A POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Gloor* for the reasons set forth above. *Supra* §§ VIII.C.2-3.

b) Dependent Claim 4

218. Dependent claim 4 recites the “method of claim 3, wherein the barcode sequences are from 4 to 20 nucleotides.” As detailed above, *Parameswaran* and

Gloor both disclose barcode sequences that are 4 to 20 nucleotides. *Supra* §§ VIII.B.2.b) (*Parameswaran*), VIII.A.2.b) (*Gloor*). A POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Gloor* for the reasons set forth above. *Supra* §§ VIII.C.2-3.

c) Dependent Claim 5

219. Claim 5 recites the “method of claim 4, wherein t [*sic*] barcode sequences are designed to have no homopolymer regions.” As detailed above, *Parameswaran* and *Gloor* both disclose barcode sequences that are designed to have no homopolymer regions. *Supra* §§ VIII.B.2.c) (*Parameswaran*), VIII.A.2.c) (*Gloor*). A POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Gloor* for the reasons set forth above. *Supra* §§ VIII.C.2-3.

d) Dependent Claim 6

220. Claim 6 recites the “method of claim 1, wherein the nucleic acid analyte of interest is on the surface of a flow cell.” Claim 6 does not require attaching the analyte to the surface of a flow cell, instead only requiring the nucleic acid analyte of interest to be “on the surface of a flow cell.” As detailed above, *Parameswaran* and *Gloor* both disclose a nucleic acid analyte of interest on the surface of a flow cell. *Supra* §§ VIII.B.2.d) (*Parameswaran*), VIII.A.2.d) (*Gloor*).

221. Further, even if the claim required *attaching* analytes to the surface of a flow cell, *Gloor* teaches this limitation. *Supra* § VIII.A.2.d) (*Gloor*). As explained above, Illumina sequencing involved attaching nucleic acid analytes to the surface of a flow cell. *See supra* §§ V.B.3-4. Further, as of December 2010, a POSA would have been motivated to use Illumina sequencing, as it was understood to be more efficient, less expensive, and faster than Roche 454 sequencers with lower error rates. *Supra* §§ VIII.C.2-3. As such, a POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Gloor* for the reasons set forth above. *Supra* §§ VIII.C.2-3.

D. Ground 4: Claims 1-6 Would Have Been Obvious Over *Parameswaran* and *Binladen*

222. As I explain above (*supra* § VIII.B), *Parameswaran* discloses each and every claim limitation of claims 1-6 of the '200 patent. As discussed below, *Binladen* also teaches many of the limitations of claims 1-6.

223. For the reasons discussed below, the combination of *Parameswaran* and *Binladen* renders obvious claims 1-6 of the '200 patent. *Supra* §§ VI.A, VI.C. A POSA would have been motivated to combine *Parameswaran* and *Binladen* to improve the accuracy of sequencing. In fact, *Parameswaran* itself studied the dual-indexing methods disclosed in *Binladen*, developed an improved dual-indexing method using longer unique identifier sequences, and demonstrated that its improved

methods led to “a more definitive assignment of samples and/or for enhanced multiplexing capabilities.” EX1004 (*Parameswaran*), 2. Further, a POSA would have had a reasonable expectation of success in combining *Parameswaran* and *Binladen* because each reference demonstrated the feasibility of using dual identifiers to exclude unwanted sequences during multiplex sequencing. As such, *Parameswaran* and *Binladen* render claims 1-6 obvious.

1. Claim 1 Would Have Been Obvious Over *Parameswaran* and *Binladen*

a) 1[pre]: “A method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction”

224. As I explain above, *Parameswaran* discloses this limitation. *Supra* § VIII.C.1.a).

225. *Binladen* also discloses this limitation. Using the Roche 454 sequencing platform, *Binladen* sequenced DNA samples extracted from thirteen species. EX1029 (*Binladen*), Abstract. Each nucleic acid analyte was dual-indexed with forward and reverse primers, the 5’ end of which additionally contained sequence tags (i.e., the “identifier sequences” in claim 1[a]). EX1029 (*Binladen*), 2 (Figure 1), 3.

226. As *Binladen* explains, “the correct association of tags and sequences [was] crucial to the approach.” EX1029 (*Binladen*), 5. Indeed, *Binladen* expressly

“discarded all sequence reads without an exact match to any of the primers used in the studies.” EX1029 (Binladen), 5. *Binladen* examined those discarded sequences in detail, and identified sequences that “lacked the 5’ tag sequence.” EX1029 (Binladen), 6. Only the remaining sequences were aligned against the reference sequences and assigned to the corresponding samples. EX1029 (Binladen), 6 (“The remaining 6307 sequences were identified through a global alignment to the 13 reference sequences. Of these, 5642 sequences (89%) diverged by no more than 1bp from one of the reference sequences, and could thus be assigned to one of the taxa analysed in the study.”).

227. As such, *Binladen* discloses a “method for validating the sequence of a nucleic acid analyte of interest in a multiplex sequencing reaction.”

b) 1[a]: “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest”

228. As I explain above, *Parameswaran* discloses this limitation. *Supra* § VIII.C.1.b).

229. *Binladen* also discloses a method of using pairs of identifier sequences. EX1029 (Binladen), 3 But unlike *Parameswaran*, *Binladen* reused pairs of identifier sequences across the nucleic acid analytes of interest. EX1029 (Binladen), 3. Identifier sequences were used for several different extractions “[t]o increase the

number of different PCR products that we could pool into the GS20-reaction beyond a single product from each of available thirteen extractions.” EX1029 (Binladen), 3.

- c) **1[b]: “at least a first identifier sequence is incorporated into a 5’ portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3’ portion of the nucleic acid analyte of interest”**

230. As I explain above, *Parameswaran* discloses this limitation. *Supra* § VIII.C.1.c).

231. *Binladen* also discloses this limitation. As depicted in **Figure 21** below, *Binladen* “label[ed] both the forward and reverse end of each PCR product” using sequence tags. EX1029 (Binladen), 3. Upon PCR, the forward and reverse tags were incorporated to the 5’- and 3’-end of each nucleic acid analyte of interest, respectively. EX1029 (Binladen), 2 (Figure 1). The forward tag incorporated to the 5’-end of the nucleic acid analyte of interest constitutes the “first identifier sequence incorporated into a 5’ portion of the nucleic acid analyte of interest.” The reverse tag incorporated to the 3’-end of the nucleic acid analyte of interest constitutes the “second identifier sequence incorporated into a 3’ portion of the nucleic acid analyte of interest.”

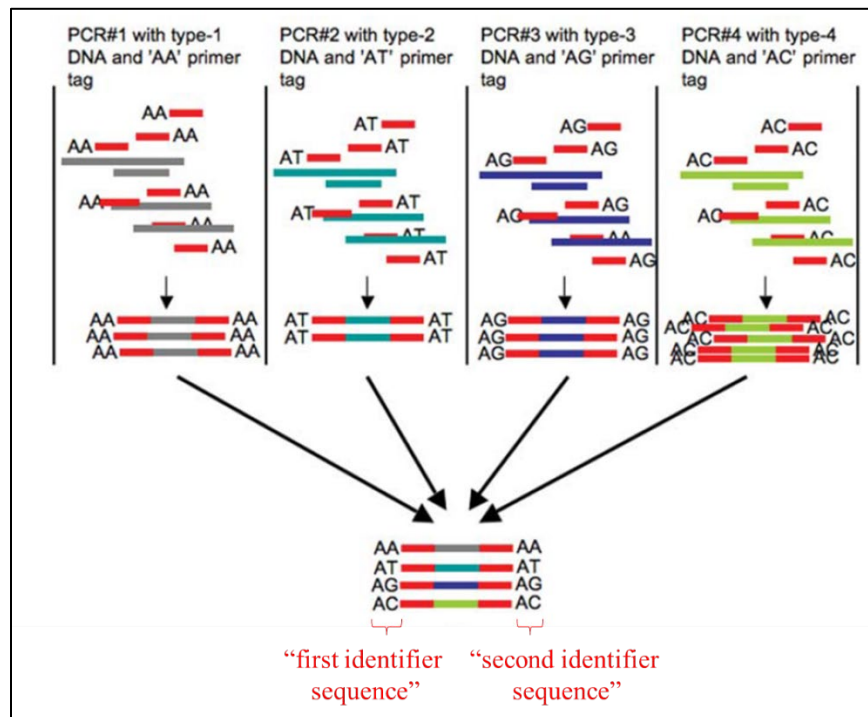


Figure 21. Application of primer tags to the GS20 sequencing-by-synthesis. (EX1029 (Binladen), 2 (Figure 1 annotated with red text and brackets))

232. As such, *Binladen* discloses that “the first identifier sequence is incorporated into a 5’ portion of the nucleic acid analyte of interest and wherein at least a second identifier sequence is incorporated into a 3’ portion of the nucleic acid analyte of interest.”

d) 1[c]: “the first and second identifier sequences have four or more nucleotides and are different”

233. As I explain above, *Parameswaran* discloses this limitation. *Supra* § VIII.C.1.d).

234. *Binladen* also discloses this limitation. As shown in *Binladen* Tables 1 and 3, *Binladen* used three pairs of first and second identifier sequences that were

four nucleotides long and were different. EX1029 (Binladen), 3 (Table 1), 7 (Table 3) (Tags 4A*, 4B*, and 4C*). Specifically, the 4A* pair included the forward tag “gcta” and the reverse tag “gtca.” EX1029 (Binladen), 7 (Table 3). The 4B* pair included the forward tag “tcag” and the reverse tag “tgac.” EX1029 (Binladen), 7 (Table 3). The 4C* pair included the forward tag “ctag” and the reverse tag “tagc.” EX1029 (Binladen), 7 (Table 3).

235. While *Binladen* also used pairs of dinucleotide tags, it explained that “miss-assignment for these [4-nucleotide tag] primers was lower than for the dinucleotide tags.” EX1029 (Binladen), 7. In assessing *Binladen*, *Parameswaran* also noted that the longer sequence identifiers allowed for “more definitive assignment.” EX1004 (Parameswaran), 2. Further, both *Parameswaran* and *Binladen* used “different” sequence identifiers, which could be beneficial for reducing sequencing error because the “same” sequence identifiers could cause analytes to be annealed together due to the perfect homology between the two sequence identifiers, creating unwanted chimeric molecules.

- e) **1[d]: “sequencing the first identifier sequence, the nucleic acid analyte of interest and the second identifier sequence”**

236. As I explain above, *Parameswaran* discloses this limitation. *Supra* § VIII.C.1.e).

237. *Binladen* also discloses this limitation. As *Binladen* explains, “approximately 50% of each PCR product will be sequenced from the orientation of the forward primer, and 50% from the orientation of the reverse primer.” EX1029 (*Binladen*), 3. As shown in **Figure 21** above, these sequence reads in combination cover the nucleic acid analyte of interest, flanked by the forward tag (i.e., the first identifier sequence in claim 1[d]) and the reverse tag (i.e., the second identifier sequence in claim 1[d]). EX1029 (*Binladen*), 2 (Figure 1); *supra* § VIII.D.1.c). *Binladen* further explains that these sequence reads were “globally aligned to the thirteen reference sequences . . . using direct and reverse complementation,” which necessarily involved sequencing and analyzing the first identifier sequence, the nucleic acid analyte of interest, and the second identifier sequence. *See* EX1029 (*Binladen*), 5.

238. Thus, *Binladen* discloses sequencing the first identifier sequence, the nucleic acid analyte of interest, and the second identifier sequence.

- f) **1[e]: “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction”**

239. As I explain above, *Parameswaran* discloses this limitation. *Supra* § VIII.C.1.f).

240. *Binladen* also discloses this limitation. As discussed above, in *Binladen*, “all sequence reads without an exact match to any of the primers used in the studies” were “discarded,” the remaining sequences with the correct forward and reverse tags were compared against the reference sequences using “direct and reverse complementation,” and nearly 90% of those sequences were “assigned to one of the taxa analysed in the study.” EX1029 (*Binladen*), 5-6. As such, *Binladen* discloses a method for “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers.”

241. *Binladen* also validated the sequence of the nucleic acid analyte of interest by excluding erroneous sequences derived from various technical errors. *Binladen* explains that each analyte generated two sets of sequence reads—each containing either the forward or reverse tag—collectively constituting the entire sequence of the analyte. EX1029, 3. Following sequencing, *Binladen* “discarded all sequence reads without an exact match to any of the primers used in the studies” (EX1029 (*Binladen*), 5), including all sequences that “lacked the 5’ tag sequence” (EX1029 (*Binladen*), 6). By excluding sequence *reads* lacking an identifier, *Binladen* necessarily excluded the sequences of analytes containing only one identifier or an incorrect pair of identifiers. As such, *Binladen* discloses “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only

one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction.”

2. A POSA Would Have Been Motivated to Combine *Parameswaran* and *Binladen*

242. A POSA in December 2010 would have been motivated to combine *Parameswaran* and *Binladen*. From *Parameswaran* and *Binladen*, a POSA would have understood that multiplex sequencing could produce unwanted erroneous sequences due to technical errors. EX1004 (*Parameswaran*), 7 (identifying “cross-contamination of template DNAs” as a potential source of erroneous sequences containing mismatched barcode motifs); EX1029 (*Binladen*), 6 (explaining that erroneous sequences like those “lack[ing] the 5’ tag sequence” could result from “errors arising during the primer synthesis process” or errors in sequencing-by-synthesis). From *Parameswaran* and *Binladen*, a POSA would have also recognized that these erroneous sequences could negatively affect the accuracy of sequencing, for example, causing “misassignment[s]” of sequences to incorrect samples. EX1004 (*Parameswaran*), Abstract; EX1029 (*Binladen*), Abstract (discussing “misassignment” rates).

243. To address this problem, a POSA would have been motivated to use the dual-indexing methods disclosed in *Parameswaran* and *Binladen*, which would allow a POSA to identify and discard erroneous sequences based on the absence of

correct identifiers. As summarized above, both *Parameswaran* and *Binladen* used dual-identifiers to differentiate and validate correct sequences obtained from various samples, versus erroneous sequences that arose through a variety of errors. EX1004 (*Parameswaran*), 8 (discussing the “remarkably low misassignment rate” achieved by its unique dual-indexing approach); EX1029 (*Binladen*), 6 (explaining that 0.4% of sequences were “miss-assigned to an incorrect identity”). Using the dual-indexing methods of *Parameswaran* and *Binladen*, a POSA would have been motivated to verify the sequences of nucleic acid analytes of interests and exclude erroneous sequences that could otherwise lead to sequence misassignments.

244. Any modification needed to adopt the dual-indexing methods of *Parameswaran* and *Binladen* would have been insignificant, as illustrated by these references. As discussed, *Binladen* used di-nucleotide and tetra-nucleotide identifier sequences, and observed that the longer sequence identifiers resulted in lower misassignment rates. EX1029 (*Binladen*), 7. *Parameswaran* characterized the dual-indexing methods disclosed in *Binladen* as “valuable,” but explained that some applications “call for a more definitive assignment of samples and/or for enhanced multiplexing capabilities.” EX1004 (*Parameswaran*), 2. To achieve that goal, *Parameswaran* used “10-base barcodes that differ from each other at multiple positions (by at least 4 nt).” EX1004 (*Parameswaran*), 2. Moreover, to further “facilitate[] resolution of ambiguities,” *Parameswaran* devised 13-nucleotide

barcode motifs by combining the 10-base barcodes and three nucleotides from the adjacent linkers. EX1004 (Parameswaran), 2, 7-8. *Parameswaran* also adopted a unique dual-indexing method where analytes from each sample were uniquely dual-indexed. These minor modifications led to a significantly lower misassignment rate. *Compare* EX1004 (Parameswaran), Abstract (reporting a < 0.005% misassignment rate) *with* EX1029 (Binladen), Abstract (reporting a < 0.4% misassignment rate). A POSA as of December 2010 would have been motivated to combine *Parameswaran* and *Binladen* to achieve the claimed validation methods.

3. A POSA Would Have Had a Reasonable Expectation of Success in Combining *Parameswaran* and *Binladen*

245. A POSA in December 2010 would have had a reasonable expectation of success in combining *Parameswaran* and *Binladen* to perform the claimed methods. As summarized above, both *Parameswaran* and *Binladen* disclosed detailed procedures for: (1) dual indexing nucleic acid analytes of interest from multiple samples; (2) sequencing them simultaneously using next-generation sequencers; (3) detecting first and second identifier sequences; and (4) validating the sequences of analytes of interest by excluding sequences containing only one identifier or incorrect pairs of identifiers. *Supra* §§ VIII.B.1, VIII.D.1. Following the teachings of *Parameswaran* and *Binladen*, a POSA would have reasonably expected that the sequences of nucleic acid analytes of interest in a multiplex sequencing

reaction would be successfully validated. In fact, a POSA reading *Parameswaran* would understand that it disclosed doing so. *Supra* §§ VI.A, VIII.B.1.

246. Based on the foregoing, claim 1 of the '200 patent would have been obvious to a POSA as of December 23, 2010, over the combination of *Parameswaran* and *Binladen*.

4. Dependent Claims 2-6 Would Have Been Obvious Over the Combination of *Parameswaran* and *Binladen*

247. For the reasons stated above in Section VIII.D.1 and below, dependent claims 2-6 also would have been obvious to a POSA over the combination of *Parameswaran* and *Binladen*.

a) Dependent Claims 2 & 3

248. Dependent claim 2 recites the “method of claim 1, wherein the identifier sequences are nucleic acid sequence tags.” Dependent claim 3 recites the “method of claim 2, wherein the identifier sequences are barcode sequences.” As detailed above, *Parameswaran* discloses a method of using identifier sequences that are nucleic acid sequence tags and barcode sequences. *Supra* § VIII.C.2.a). *Binladen* likewise discloses identifier sequences that are nucleic acid sequence tags and barcode sequences. EX1029 (*Binladen*), 3. For instance, *Binladen* disclosed “dinucleotide tags” and “tetranucleotide tags,” each of which was made of nucleic acids, was incorporated into the nucleic acid analytes of interests, and functioned as

a barcode sequence. EX1029 (Binladen), 3, 6-7 (Table 3). A POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Binladen* for the reasons set forth above. *Supra* §§ VIII.D.2-3.

b) Dependent Claim 4

249. Dependent claim 4 recites the “method of claim 3, wherein the barcode sequences are from 4 to 20 nucleotides.” As detailed above, *Parameswaran* discloses barcode sequences that are 4 to 20 nucleotides. *Supra* § VIII.C.2.b). As illustrated in Tables 1 and 3, *Binladen* likewise discloses the use of barcode sequences that are 4 nucleotides in length. EX1029 (Binladen), 3 (Table 1), 7 (Table 3). Specifically, as shown in Table 3, each of the 4A*, 4B*, and 4C* pairs included two 4-nucleotide tags. EX1029 (Binladen), 7 (Table 3). A POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Binladen* for the reasons set forth above. *Supra* §§ VIII.D.2-3.

c) Dependent Claim 5

250. Claim 5 recites the “method of claim 4, wherein t [sic] barcode sequences are designed to have no homopolymer regions.” As detailed above, *Parameswaran* discloses barcode sequences that were designed to have no homopolymer regions. *Supra* § VIII.C.2.c). As illustrated in Tables 1 and 3, *Binladen* likewise discloses 4-nucleotide identifier sequences having no homopolymer regions. EX1029 (Binladen), 3 (Table 1), 7 (Table 3). A POSA would

have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Binladen* for the reasons set forth above. *Supra* §§ VIII.D.2-3.

d) Dependent Claim 6

251. Claim 6 recites the “method of claim 1, wherein the nucleic acid analyte of interest is on the surface of a flow cell.” Claim 6 does not require attaching the analyte to the surface of a flow cell, instead only requiring the nucleic acid analyte of interest to be “on the surface of a flow cell.” As detailed above, *Parameswaran*, which used Roche 454 sequencing, disclosed a nucleic acid analyte of interest on the surface of a flow cell. *Supra* § VIII.C.2.d). *Binladen* likewise used Roche 454 sequencing, which similarly involved placing nucleic acid analytes of interest on the surface of a flow cell. EX1028 (Roche 454 Manual), 6-8. A POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Binladen* for the reasons set forth above. *Supra* §§ VIII.D.2-3.

E. Ground 5: Claims 6 Would Have Been Obvious Over *Parameswaran*, *Binladen*, and *Bentley*

252. Claim 6 recites the “method of claim 1, wherein the nucleic acid analyte of interest is on the surface of a flow cell.” As discussed above, claim 6 does not require attaching the analyte to the surface of a flow cell, instead only requiring the nucleic acid analyte of interest to be “on the surface of a flow cell.”

Supra §§ VIII.C.4.e), VIII.D.4.e). *Parameswaran* and *Binladen* disclose this limitation. *Supra* § VIII.D.4.e).

253. Even if the claim required *attaching* analytes to the surface of a flow cell, *Bentley* discloses this limitation. *Supra* § VI.D. *Bentley* discloses Illumina sequencing, where the nucleic acid analytes of interest were sequenced on the surface of a flow cell. EX1006 (*Bentley*), 1-2. Specifically, *Bentley* explains that “[w]e generated high-density single-molecule arrays of genomic DNA fragments attached to the surface of the reaction chamber (the flow cell) and used isothermal ‘bridging’ amplification to form DNA ‘clusters’ from each fragment.” EX1006 (*Bentley*), 1-2 (“we generated 1-2 Gb of high-quality purity filtered sequence per flow cell”).

254. A POSA would have been motivated to combine *Parameswaran*, *Binladen*, and *Bentley* with a reasonable expectation of success. As discussed above, as of December 2010, a POSA would have been motivated to use the dual barcoding methods of *Parameswaran* and *Binladen* with Illumina sequencing as disclosed in *Bentley*, as Illumina sequencing was understood to be more efficient, less expensive, and faster than Roche 454 sequencing with lower error rates. *Supra* § VIII.D.2. A POSA would have been further motivated to use the dual indexing methods of *Parameswaran* and *Binladen* on an Illumina sequencer to reduce error rates by

identifying and excluding sequences with only one identifier or incorrect pairs of identifiers. *Supra* § VIII.D.2.

255. Further, a POSA would have had a reasonable expectation of success in combining *Parameswaran*, *Binladen*, and *Bentley*. *Supra* § VIII.D.3. *Parameswaran* explains that its dual-indexing method would be compatible “with other sequencing platforms,” and identified Illumina sequencing as one of the “major high-throughput sequencing platforms.” EX1004 (*Parameswaran*), 1, 2, 8. A POSA would have reasonably expected that this dual-indexing method could be used on the Illumina sequencing platform used in *Bentley* by simply swapping Roche 454 adapters for Illumina-specific adapters and making minor, routine modifications to the sequencing workflow. EX1015 (*Meyer 2010*), 1; *supra* §§ VIII.C.3, VIII.D.3. A POSA would have understood that these modifications do not require altering the fundamental aspect of the dual-indexing methods disclosed in *Parameswaran* and *Binladen*. In fact, before December 2010, researchers had already performed dual-index sequencing on both Roche 454 and Illumina sequencers using the same set of barcode pairs for both platforms. *See generally* EX1038 (*Maughan*).

256. The relatively short read lengths reported in *Bentley* do not change my conclusion. *Bentley*, which was published in 2008, discloses read lengths of 35-bases. EX1006 (*Bentley*), Abstract. A POSA as of December 2010, however, would have known that the read length of Illumina sequencing had significantly

increased, surpassing the read lengths required for *Binladen*'s and *Parameswaran*'s dual-index sequencing. Indeed, by December 2010, Illumina sequencers were capable of "high-quality paired end reads of 150 base pairs (bp)." EX1027 (July 2010 Press Release), 1. This is longer than the "full sequence" of 133-141 bp products (including primers) sequenced in *Binladen* and is nearly two-fold longer than the read lengths required for each template nucleic acid in *Parameswaran*. See EX1029 (*Binladen*), 3; EX1004 (*Parameswaran*), 3 (Figure 1) (the read lengths between 71 and 81 nucleotides: two barcodes (10 nucleotides each), two cloning linkers (16-17 nucleotides each), and one small RNA insert (19-27 nucleotides)). Thus, a POSA in December 2010 would have known that Illumina sequencing was available and appropriate for *Parameswaran*'s or *Binladen*'s dual-indexing method and would have had a reasonable expectation of success.

257. Thus, a POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran*, *Binladen*, and *Bentley* for the reasons set forth above. *Supra* §§ VIII.D.2-3, VIII.C.2-3.

F. No Unexpected Results or Other Evidence of Nonobviousness

258. I did not identify any evidence, such as unexpected results, that would support the nonobviousness of the claims of the '200 patent. I reserve the right to address any such evidence if it is identified by Molecular Loop or its experts.

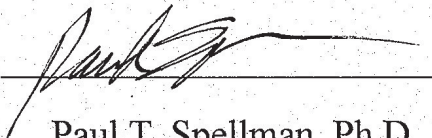
259. The '200 patent asserts that "FIG. 3 also provides data showing that methods of the invention dramatically reduce false positives in a multiplex sequencing reaction." EX1001 ('200 patent), 6:53-55. But the '200 patent only compared single-indexing versus dual-indexing. *Supra* § VII.C. A POSA would have already known that previously disclosed methods of dual-indexing were superior to single indexing for validating the sequence of nucleic acid analytes of interest. EX1004 (Parameswaran), 7; EX1005 (Gloor), 1-2; EX1029 (Binladen), 5. Further, as the '200 patent acknowledges, "[n]umerous techniques are known for detecting sequence and for identifying barcodes...[h]owever, the exact means for detecting and compiling sequence data does not affect the function of the invention." EX1001 ('200 patent), 3:44-48.

IX. CONCLUSION

260. For the reasons detailed above, it is my opinion that claims 1-6 of the '200 patent are anticipated and would have been obvious to a POSA as of the effective filing date.

261. 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 knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001. I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

DATE: July 1, 2024

By: 
Paul T. Spellman, Ph.D.

Appendix A

Paul T. Spellman, Ph.D.

Curriculum Vitae

Updated May 31, 2024

BUSINESS INFORMATION

Business Address: **University of California Los Angeles**
 Department of Medicine
 Department of Human Genetics
 David Geffen School of Medicine

 10833 Le Conte Ave
 Los Angeles, CA 90095

E-Mail Address: **PSpellman@mendnet.ucla.edu**

EDUCATION

S.B., Biology, Massachusetts Institute of Technology, 1995
Ph.D., Genetics, Stanford University School of Medicine, 2000

PROFESSIONAL EXPERIENCE

2023 –	Professor of Medicine and Human Genetics, UCLA David Geffen School of Medicine
2013 – 2023	Professor with tenure, Oregon Health & Sciences University, Department of Molecular and Medical Genetics
2017 – 2023	Program Leader for Quantitative Oncology, Knight Cancer Institute
2017 – 2023	Co-Director, CEDAR Center
2018 – 2020	Interim Director, Program in Computational Biology
2013 – 2014	Interim Director, Program in Computational Biology
2011 – 2013	Associate Professor, Oregon Health Sciences University, Department of Molecular and Medical Genetics
2011	Special Assistant to the Deputy Director of the National Cancer Institute
2007 – 2011	Staff Scientist, Lawrence Berkeley Lab, Life Science Division
2003 – 2007	Scientist, Lawrence Berkeley Lab, Life Science Division
2000 – 2003	Post-doctoral Fellow, UC Berkeley, Department of Molecular and Cellular Biology, Gerald M. Rubin Advisor

PROFESSIONAL ACTIVITIES

Granting Agency Review Work:

2002	Ad Hoc member, NSF, GEN-EN panel
2005	Ad Hoc member, EPA, Center for Bioinformatics Panel

2006 – 2007 Panel member, NIH/NHLBI, Exploratory Program in Systems Biology
 2012 SAIC panel review – SEG DCC Competition
 2012 NIH Innovator Awards
 2012 Susan G. Komen Postdoctoral Awards Reviewer
 2013 Ad Hoc member Cancer Genetics study section
 2014 – 2018 Member Cancer Genetics study section
 2017 CDMRP Breast Cancer study section member
 2018 CDMRP Breast Cancer study section member
 2019 Jackson Lab Cancer Center Site Visit Reviewer
 2019 University of Illinois Champaign Urbana Cancer Center Site Visit Reviewer
 2019 NIH New Innovator DP2 Review
 2020 Herbert Irving Comprehensive Cancer Center (Columbia) Site Visit Reviewer
 2020 NCI Answers to Provocative Questions Review
 2020 NIH New Innovator DP2 Review
 2020 NCI SPORE Review
 2021 University of New Mexico Cancer Center Site Visit Reviewer
 2021 Fox Chase Cancer Center Site Visit Reviewer
 2022 NCI SPORE Review
 2022 Project Reviewer BreastCancerNow
 2022 V Foundation
 2023 NCI MetNet Review
 2023 Canadian Excellence Research Chairs (CERC) Competition 3 midterm review
 2024 Indiana University Simons Comprehensive Cancer Center Site Visit Reviewer

Conference Chair:

2002 Program planning committee, Standards and Ontologies for Functional Genomics. Hinxton, UK. November 17-20, 2002
 2006 Conference co-organizer MSRI Mathematical Systems Biology Conference, Berkeley, CA
 2009 Conference co-organizer MSRI Mathematical Systems Biology Conference, Berkeley, CA
 2009 Conference Chair, MGED12 Phoenix, AZ
 2010 Conference co-organizer, RECOMB Cancer Computational Biology, Oslo, Norway
 2011 Scientific Session Organizer, International Congress of Human Genetics, Montreal, Canada

National Panels

2001 National Academy of Sciences, Computing Panel, Title “Development of Data Communication and Storage Formats for Microarray Data”
 2003 National Academy of Sciences, Committee on Emerging Issues & Data on Environmental Contaminants, September, Seattle, WA, Title: “MAGE: A Markup Language for MicroarrayData”

Professional Societies

2002 – 2006 Treasurer, Microarray Gene Expression Data Society

2002 – 2010 Board member, Microarray Gene Expression Data Society
 2007 – 2011 International Cancer Genomics Consortium, Data Coordination and Management Working Group
 2007 – 2008 International Cancer Genomics Consortium, Scientific Planning Committee

External Advisory Boards

2008 Welcome Trust Cancer Genome Sequencing Project External Advisory Board
 2012 – 2015 Canadian Prostate Cancer Network SAB
 2012 National Cancer Advisory Board, Center for Cancer Genomics Working Group
 2013 – 2014 National Cancer Advisory Board, ad hoc Information Technology Working Group
 2013 National Cancer Institute Scientific Program Leaders/Office of Director Staff Retreat
 2018 Triple Negative Breast Cancer (TNBC) External Advisory Board, Indiana University Precision Health Initiative

National/International Project Advisory Boards

2008 – 2010 Integrative Cancer Biology Program (NCI) Data Sharing Working Group
 2009 – International Cancer Genomics Consortium, Scientific Steering Committee
 2009 – 2015 International Cancer Genomics Consortium, Bioinformatics Analyses Working Group Chair
 2010 International Cancer Genomics Consortium, International Planning committee
 2011 – 2015 International Cancer Genomics Consortium, Executive Steering Committee
 2011 National Cancer Institute, 2011 IPA at request of NCI Director Harold Varmus
 2011 TCGA Executive Steering Committee, NIH
 2011 TCGA Project Team, NIH

 2011 Faculty review of Junior Faculty candidate Katholieke Universiteit Leuven

Editorial and Ad Hoc Review Activities:

Regular reviewer for many journals including Bioinformatics, NAR, Nature Communications, PLoS ONE, Cell Reports, Genome Research, Genes, etc...

Institutional Review Activities:

2011 – OHSU Genomics Advisory Committee
 2011 – 2012 OHSU Computational Biology Task Force
 2011 – 2012 OCSSB Biology Faculty Search Committee
 2011 – 2012 OCSSB Bioinformatics/Computational Biology Faculty Search Committee
 2012 Hematology and Oncology Division Chair Search Committee
 2012 – 2013 DNA Service Committee, Tom Keller
 2012 – 2016 Task Force 3, hosted by Molecular & Medical Genetics Chair, Dr. Susan Hayflick
 2012 Director of Computational Biology Search Committee
 2012 – 2016 Intel – OHSU Project Working Group
 2012 OCSSB Pilot Project Review Committee
 2018 – 2019 Center for Spoken Language Understanding Review Committee

Institutional Committees:

2008 – 2010 LBL Computing task force
 2011 – 2019 PMCB Admissions Committee
 2012 – 2018 OHSU Research Committee
 2016 – 2018 Molecular and Medical Genetics Diversity Committee
 2017 – OHSU School of Medicine Diversity Affairs Committee
 2018 – Medical Scientist Training Program MD/PhD Dual Degree Program Committee
 2021 OHSU Faculty Compensation Committee

Community Service:

Regular external reviewer for P&T committees including, Harvard, Ontario Institute for Cancer Research, Boston University, Stanford, Harvard School of Medicine, The Ohio State University, Koc University, University of British Columbia

HONORS & AWARDS

2018 Penny and Phil Knight Endowed Professorship in Cancer Research Innovation
 2020 AACR Team Science Award: THE CANCER GENOME ATLAS (TCGA) PILOT PROJECT

RESEARCH GRANTS AND FELLOWSHIPS RECEIVED

Grants and Contracts:

Systems-Based Predictions of Responses to Cancer Therapy. NIH/NCI U54 CA112970. \$250,000 Role: Project 1 Leader. (PI Gray, Joe W.) 50% effort. 09/30/04-08/31/10

NIH/NHGRI P41 G003619-01. Role: Project Leader. (PI Catherine Ball). 20% effort. 8/1/2005-7/31/2008

Development of cost effective sequence-based technologies to identify genome. Department of Energy DE-AC02-05CH11231. Role: Co-PI (PI Pennachio, Len) 5% effort. 10/1/05-9/30/07

Development of RPPA data modules for caArray. National Cancer Institute Y1-CO-6120-01. 135,000. Role: PI. 10% effort. 09/19/06-07/31/09

Univ. of Calif., San Francisco (NIH/NCI Prime) U24 CA126477 Role: Scientist (PI Fisher, Susan J.) 20% effort. 09/26/06-08/31/11

The Berkeley Cancer Genome Center NIH/NCI. U24 CA126551 \$1,250,000 CoPI (PI Gray, Joe W.) 30% effort 09/28/06-08/31/10

SPORE in Breast Cancer Univ. of Calif., San Francisco (NIH/NCI Prime) P50 CA058207. \$53,820. Role: Co-investigator (PI van't Veer, Laura). 8% effort. 09/30/07-11/30/12

TCGA Data Analysis Center at Berkeley NIH/NCI U24 CA143799. \$537,280. Role: PI 20% effort. 09/30/09-07/31/14

Model-based Predictions of Responses to RTK Pathway Therapies. NIH/NCI. U54 CA112970 \$520,700 Role: Project 1 PI and Data & Model Repository Core PI. (PI Gray, Joe W.) 17% effort. 09/1/10-02/28/15.

Enhancing the Breadth and Efficacy of Therapeutic Vaccines for Breast Cancer. DoD BC100597P1. \$383,443. Role: PI 30% effort 9/30/11-9/29/15

Complications During Long-duration Space Missions NSBRI/Baylor CA04201. \$232,073. Role CoI (PI: Lindner, Jonathan) 0.36 calendar months 06/01/15-05/31/17.

Extrinsic Perturbations of Cell Physiology and Associated Regulatory Networks NIH/NHGRI U54HG008100. \$1,396,812. Role CoI (PI: Gray). 0.6 calendar months 09/10/14-06/30/20

Intratumor heterogeneity underlying treatment resistance in HER2+ breast tumors NIH/NCI U01CA195469). \$165,999. (PI: Gray, Michor, Polyak, Spellman – Multip-PI) 0.84 calendar months 06/01/15-05/31/20

Control of genetic and epigenetic instabilities by lincRNA genes NIH/NIHMS R01GM114162 \$192,500. CoI (PI:Thayer) 0.6 calendar months 09/17/15-06/30/19

Integrative approach to heterogeneity in breast metastasis. UCSF/NIH U01CA199315 \$166,414. (PI: Werb and Spellman) 1.8 calendar months 06/01/16-05/31/21

OHSU Center for Specialized Data Analysis as part of the GDAN NIH/NCI U24CA210957). \$273,900. (PI: Spellman / Margolin) – Multi-PI 2.4 calendar months 09/13/16-08/31/21

Molecular Changes Associated with Initial Steps of Breast Cancer Metastasis for Risk Stratification and Therapy Development DOD/BCRP BC151431P1 \$148,059. (PI: Spellman). 0.48 calendar months 09/30/16-08/31/19

Quantitative Oncology NIH/NCI 5P30CA069533. \$190,837 total cost. (PI Spellman). 0.9 calendar months 07/01/17-06/30/22

Evaluation of Population Based Testing for HBOC and Lynch Syndromes \$4,627,812 direct costs. (MPI Spellman / Shannon). 1.2 cal. 06/10/2020-05/31/2025

OHSU Center for Specialized Data Analysis as part of the GDAN U24CA264007. \$369,600/year total cost. PI: Spellman / Ellrott – Multi-PI 2.4 calendar months 09/01/21-08/31/26

Examining the Impact of Structural Racism on African American Non-Small Cell Lung Cancer Mutational Signatures and Outcomes ~100,000 direct/year 1R37CA284040-01A1 (PI Erhunmwunsee) 1.2 calendar months 05/10/24-05/09/31

Other Support (Include Title, Source, PI, Amount Period, and % Effort)

Molecular Predictors of Drug Response in Breast Cancer SmithKline Beecham Corporation
Role: Scientist (Gray, Joe W.) 5% effort 01/27/06-01/26/10

MIPS Analysis The Buck Institute Role: PI (Spellman, Paul T.) 2% effort. 05/01/07-04/30/08

An Integrated Approach to Targeting Breast Cancer Molecular Subtypes and Their 'Resistance' Phenotypes University of California at Los Angeles (American Association for Cancer Research/Stand Up To Cancer, Prime) SU2C-AACR-DT0409. \$915,164 Role: Co-Investigator 10% effort (PI: Slamon, Dennis J. and Gray, Joe W.) 10/01/09 – 09/30/12

The development of the ovarian cancer genome Anna Fuller Fund \$100,000. Role: PI (Spellman, Paul T.) 0% effort 03/01/12-02/28/14

Identification of breast cancer-specific antigen targets for development of anti-cancer vaccines National Philanthropic Trust National Breast Cancer Coalition – Artemis Project. \$129,714. Role: PI 10% effort (Spellman, Paul T.) 8/1/12-7/31/13

Restoring hormone sensitivity of late recurrences in breast cancer patients. Promise Grant (O'Malley, Bert W) Susan G. Komen Foundation / Baylor College of Medicine. \$134,838. Role: Co-investigator (Gray) 2% effort 09/13/12-09/12/17

Accelerator-Rich Architectures with Applications to Healthcare Intel/UCLA. \$25,974 (PI: Cong) 0.6 calendar months 01/01/14-08/31/17

ctDNA profiling to detect early recurrence Prospect Creek Foundation \$238,096. (PI: Gray) 0.6 calendar months 10/01/15-09/30/17

ctDNA profiling to detect early recurrence Prospect Creek Foundation \$158,730. (Gray) 0.6 cal. mos. 10/01/15-09/30/18

Widespread Testing for Cancer Risk Syndromes Knight Cancer Institute \$311,373. (Spellman). 09/01/17-08/31/18

LECTURES AND PRESENTATIONS

International and National

1. First Virtual Conference on Genomics and Bioinformatics, October 15-16 2001, Fargo, ND. "Development of MAGE-ML"
2. 6th annual Advancing Pathology Informatics, Imaging, and the Internet conference, October 4-6 2002, Pittsburgh, PA. "MAML: The Gene Expression Microarray Experience with Data Exchange Standards"
3. Chips to Hits, October 27-31, 2002, Philadelphia, PA. "Design and Implementation of MAGE-ML"
4. Gordon Research Conference on Aging workshop, March 14-15, 2003, Ventura, CA, "Challenges with Microarray Data"

5. Ontario Microarray Network Technology Symposium, November 2003, Toronto ON, Canada. "MGED & MAGE"
6. Conference speaker, AACR Annual Meeting, April 12-16, 2007, Los Angeles, CA. "MIP Based Allele Copy Number Measurements"
7. International Congress on Radiation Research, July 8-12, 2007, San Francisco, CA. "Systems Biology of Cancer"
8. International Cancer Genomics Consortium Planning Meeting, October 1-2, 2007, Toronto, ON, Canada. "Data Integration, Analysis, and Visualization"
9. Society for Neuro-oncology, November 16-18, 2007, Dallas, TX. "A TCGA GBM Data Primer"
10. Biology of GBM TCGA Workshop, Mar 24-25, 2008, Bethesda, MD. "An update on GBM Analyses"
11. European Association of Cancer Research Biannual Meeting, July 5-8, 2008, Lyon, France. "An In Vitro Systems Approach to Predicting and Understanding Clinical Responses to Molecularly Targeted Therapeutics"
12. RECOMB -- Computation Cancer Biology, September 19-20, 2008, San Diego, CA. "Systems Biology of Cancer at LBL"
13. IBC Assay and Screening Advancements, September 24-26, 2008, San Diego, CA. "Genomic Approaches to Understanding Cancer Biology and Treatment"
14. Symposium on Clinical Care in Oncology, October 17-19, 2008, Huntington Beach, CA. "Genomic Approaches to Understanding Cancer Biology and Treatment"
15. HCCS 2008, November 12-15, 2008, Singapore. "Genomic Approaches to Understanding Cancer Biology and Treatment"
16. 1st ICGC Scientific Planning Committee Meeting, November 15-17, 2008, Washington DC. "State of Ovarian Omics"
17. Mouse Models of Human Cancer Consortium Steering Committee meeting, January 11-12, 2009, San Francisco, CA. "Interpretations of Cancer Genomes Through Integrative Genomics"
18. ICBP/CISNET meeting, March 24-25, 2009, Washington, DC
19. Keystone: Targeted Cancer Therapies, March 27-31, 2009, Whistler, BC, Canada, "Omic approaches to predictive markers"
20. AACR, April 21, 2009, Denver CO. NCI special session on TCGA. "A Preliminary Look at Data from the Serous Ovarian Carcinoma Project"
21. 3rd Int'l Symposium on Cancer Metastasis and the Lymphovascular System: Basis for Rational Therapy, May 6-9, 2009, San Francisco, CA. "Omic approaches to predictive markers"
22. WNAAR/IMS, June 14-17, 2009, Portland, OR. "Next-gen Sequencing and RNAseq Data: An Overview"
23. Second ICGC Scientific Planning Committee Meeting, June 22-24, 2009, Hinxton, UK. "Update on TCGA"
24. Clinical Proteomic Technologies Strategy Workshop: Implementation of a New Cancer Biomarker Development Process, September 8-9, 2009, Bethesda, MD. "Subtypes and Mutations"
25. University of Southern California, November 12, 2009, Los Angeles, CA. Department of Molecular and Computational Biology. "A Look at the TCGA Pilot Project"

26. Integrating Knowledge of Inherited Genetic Variation and Acquired Somatic Alterations in Cancer Tissues: Promising Scientific Opportunities, December 1-2, 2009, Bethesda, MD. "TCGA Ovarian Cancer Project"
27. 3rd Scientific Workshop of the International Cancer Genome Consortium (ICGC), March 21-23, 2010, Madrid, Spain. "Large scale genomic studies of serous ovarian carcinoma"
28. Intelligent Systems for Molecular Biology 2010. Boston MA July 11, 2010, "Pancreas and Kidney Projects within TCGA"
29. Targeting Cancer Genomics and Cancer Mutations, Burlingame CA, August 2-4, 2010, "Molecular Subtype Specific Responses to Anticancer Compounds in Breast Cancer"
30. AMATA 2010, September 15, 2010, Hobart Australia. Keynote Speaker Title "Systems Biology of Cancer"
31. 4th AACR Conference on Molecular Diagnostics, Denver CO, September 27-30, 2010, "Impact of Systems Biology on Molecular Diagnostics"
32. Bridging the Gap in Biomedical Genetics: Integrating Bioinformatics and Medical Informatics, October 27-29, 2010, Hinxton UK, "The Ovarian Cancer Genome"
33. Walter and Eliza Hall Institute, November 23, 2010, Melbourne, Australia, "Systems Biology of Cancer"
34. 4th Scientific Workshop of the International Cancer Genome Consortium (ICGC), December 1-3, 2010, Couran Cove, Australia, "The Ovarian Cancer Genome"
35. Pacific Symposium on Biocomputing, January 3-7, 2011, Kona HI, "A Challenge in Network Reconstruction – **The Cancer Genome Atlas**"
36. AACR-NCI Conference on Systems Biology: Confronting the Complexity of Cancer, February 27- March 2, 2011, La Jolla CA, "**The Cancer Genome Atlas**"
37. AACR-NCI Conference on Systems Biology: Confronting the Complexity of Cancer, February 27- March 2, 2011 La Jolla CA, "Therapeutic Insights from Integrated Cancer Genomics – Breast Cancer as an Example"
38. EDRN, March 8-10, 2011, Los Angeles, CA, "**The Cancer Genome Atlas**"
39. Mouse Models of Human Cancer Consortium, June 21-2, 2011, Bethesda MD, "MMHCC Genetics Group Mouse TCGA"
40. NCI Workshop on Criteria for Use of Omics-Based Predictors in Clinical Trials, June 23-4, 2011, Bethesda MD, "Sample Reuse in Large Scale Gene Expression Studies"
41. Spanish Association for Cancer Research XIII National Conference (ASEICA), September 21-3, 2011, Salamanca, Spain, "**The Cancer Genome Atlas**"
42. EORTC-NCI-ASCO Annual Meeting on "Molecular Markers in Cancer", October 25-7, 2011, Brussels Belgium, "TCGA-Success Story in Ovarian Cancer"
43. US HUPO, San Francisco, CA, March 4-7, 2012, "Lessons from TCGA"
44. Frontiers in Cancer, Karolinska, Stockholm Sweden, March 8-9, 2012, "Systems Approaches to Cancer"
45. Biology of Genomes, Cold Spring Harbor Laboratory, May 8-12, 2012, "The Evolving Cancer Genome"
46. GAME-ON, Washington DC, July 27-29, 2012, "Systems Approaches to Cancer Biology"
47. Netherland Cancer Institute, Amsterdam, December 12, 2012, "Cancer Genomics"
48. Komen Foundation Global Summit on Women's Health, February 4, 2013, Panelist on use of technology to fight breast cancer

49. National Breast Cancer Coalition, Project Artemis Meeting, March 9-11, 2013, “Antigens for Preventative Vaccine Development”
50. American Association for Cancer Research Annual Meeting, Washington DC, April 6-10, 2013, “Reconstruction of Distinct Signaling Networks in Breast Cancer Cell Lines”
51. 10th International Workshop on Advanced Genomics, Tokyo, Japan, May 21-23, 2013, “Application of Genomic Technologies to the Understanding of Cancer”
52. Stanford University Cancer Systems Biology Seminar Series, August 23, 2013 “The Thousand Faces of Breast Cancer”
53. American Society for Biochemistry and Molecular Biology Annual Meeting, April 27, 2014, “Reconstructing Signaling Networks in Cancer”
54. Early Detection Research Network Meeting, September 10, 2014, “Impact Of The TCGA In Ovarian Cancer”
55. AACR Special Meeting on Computational and Systems Biology, February 7, 2015, “Reconstructing Signaling Networks in Cancer”
56. MSKCC Invited Seminar Series, Systems Biology Seminar, April 8, 2015, “Two Topics in Cancer”
57. 6th Pac Rim Breast & Prostate Meeting, June 23, 2015, “Chronological Dating of Chromosome Duplications in Tumor Development
58. Computation Cancer Biology, Simons Institute, Berkeley CA, February 3, 2016, “Evolutionary Processes, Constraints and Drivers in Kidney (and other?) Cancers”
59. University of Texas Health Sciences Department of Integrative Biology Seminar Series, March 21, 2016, “Genomic Approaches to Studying Cancer”
60. Columbia University Department of Systems Biology Seminar Series, April 27, 2016, “Genomic Approaches to Studying Cancer”
61. OHSU/CRUK Early Detection Conferences, June 22, 2016, “Genomic Approaches it Studying Early Cancer Etiology”
62. International Association of Breast Cancer Research Komen Public Seminar, August 2016, “Blood Based Screening for Breast Cancer Early Detection”
63. International Association of Breast Cancer Research, August 2, 2016, “Chronological Dating of Chromosome Duplications in Tumor Development”
64. CRUK Manchester Conference on Early Detection, September 28, 2016, “Chronological Dating of Chromosome Duplications in Tumor Development”
65. SWOG Radiation Symposium, April 29, 2017, “A Proposed Study to Evaluate the Benefit of Radiation in Mastectomy Treated Breast Cancer”
66. Case Western Reserve Seminar Series, May 18, 2017, “Proteomic and Genomic Approaches to Studying Cancer”
67. Jackson Laboratory Course, August 13, 2017, “Circulating-Tumor DNA”
68. UCSF Breast Oncology Program, October 25, 2017, “Breast Cancer Risk and Monitoring”
69. Stanford Genetics 65th Anniversary, July 26-27, 2018, “Walk with me: Making a Team Scientist”
70. AACR Annual Meeting, March 30, 2019, “ Key Questions in ctDNA Analysis”
71. ASHG, November 2, 2023, “HOP POP”
72. DREAM RECOMB Keynote, November 30, 2023, “Constraints on Tumor Development”
73. PMWC, January 26, 2024, “HOP POP”

Regional and Local

1. UCSF Cancer Genetics Program, September 8, 2009. "Cancer Genomes: what will we find"
2. UCSF Breast Oncology Program, UCSF San Francisco, CA January 29, 2010. "A Look at the TCGA Project GBM, Ovarian, & Breast"
3. Illumina Users Meeting, March 7-9, 2010. Napa CA. "Large scale genomic studies of serous ovarian carcinoma"
4. Biomedical Engineering Weekly Seminar, June 1, 2011, "Tackling the Complexity of Cancer Using Systems Biology"
5. Knight Cancer Institute, Weekly Seminar, October 4, 2011, "Systems approaches to understanding cancer"
6. Reed College, Weekly Seminar, April 4, 2014, "Reconstructing Signaling Networks in Cancer"
7. DNA Damage Symposium, October 24, 2014, "Plasma-based Sequencing to Characterize Cancers and Recurrences"
8. Portland State University Systems Science Seminar Series, December 5, 2014, "Reconstructing Signaling Networks in Cancer"
9. OHSU Molecular and Medical Genetics Grand Rounds, April 16 2015, "cfDNA and ctDNA"
10. Center for Domain Specific Computing, October 24, 2016, "Chronological Dating of Chromosome Duplications in Tumor Development"
11. OHSU Molecular and Medical Genetics Grand Rounds, November 3, 2016, "cfDNA and ctDNA"
12. OHSU Medical Oncology Fellow Seminar Series, February 10, 2017, "Genetic Constraints on Tumor Evolution"

TEACHING ACTIVITIES

Teaching

Winter 2012	CON663, One lecture, Systems Biology
Winter 2012	CANB610, Eight classes, guided journal club format
Spring 2012	CELL616, One lecture, Predictors of Response in Breast and Ovarian Cancer
Spring 2012	CON668, One lecture, 2 nd Generation Sequencing Technologies and Applications
Fall 2012	CON650, One lecture, Practice and Ethics of Science
Winter 2013	Cell Structure Function, one lecture, The Human Genome
Winter 2013	CON663, One lecture, Systems Biology
Spring 2013	CELL616, One lecture, Predictors of Response in Breast and Ovarian Cancer
Spring 2013	CON665, One lecture, Cancer Genetics
Spring 2014	Course Director CELL616
Spring 2014	CELL616, One lecture, Predictors of Response in Breast and Ovarian Cancer
Fall 2014	CONJ670, One lecture, DNA sequencing technologies
Winter 2015	CON663, One lecture and one journal club, Systems Biology
Winter 2015	Course Director CONJ671
Winter 2015	CONJ671, One lecture DNA Sequence analysis
Spring 2015	Course Director CELL616
Spring 2016	Course Director CELL616

Fall 2017	Fundamentals, four lectures
Winter 2017	Course Director CONJ671
Fall 2018	Fundamentals, three lectures
Fall 2019	Fundamentals, three lectures
Spring 2020	CONJ665 Development, Differentiation and Disease
Fall 2020	Fundamentals, GWAS and Gene Expression I & II
Fall 2021	Fundamentals, GWAS, Population Genetics, and Gene Expression I & II
Winter 2024	Bioinformatics

Ad hoc

1. UC Davis, Bioinformatics Summer Course. August 2-7, 2001. Title “Clustering Gene Expression Data”
2. UC Berkeley Special Seminar Series in Statistical Genomics. September 12, 2002. Title “Domains of Similar Expression in *Drosophila melanogaster*”
3. UC Berkeley Special Seminar Series in Statistical Genomics. November 2, 2006. Title “Systematic Approaches for Characterizing Cancer”.
4. ACED February 15, 2022 Title “Cancer Genomics”

Education Grants and Contracts

2010 – 2011 Postdoctoral Fellowship Advisor CHIR, Obi Griffith
 2010 – 2011 Postdoctoral Fellowship Advisor – BAEF, Anneleen Daemen
 2018 – 2021 PhD fellowship advisor – F99/R00, Michael Heskett

Workshop Presentations Invited guest lectures in courses or educational seminars (* items had both research and didactic content and are in both sections of CV)

1. First Virtual Conference on Genomics and Bioinformatics, October 15-16, 2001, Fargo, ND. “Development of MAGE-ML”*
2. 6th annual Advancing Pathology Informatics, Imaging, and the Internet conference, October 4-6 2002, Pittsburgh, PA. “MAML: The Gene Expression Microarray Experience with Data Exchange Standards”*
3. Chips to Hits, October 27-31, 2002, Philadelphia, PA. “Design and Implementation of MAGE-ML”*
4. Gordon Research Conference on Aging workshop, March 14-15, 2003 Ventura, CA, “Challenges with Microarray Data”*
5. Ontario Microarray Network Technology Symposium, November 2003, Toronto ON, Canada. “MGED & MAGE”*
6. Workshop Speaker, Emerging Genomic Technologies and Data Integration Problems, December 2004, Ohio State University, Columbus, OH. “A System for Studying Cancer”
7. International Cancer Genomics Consortium Planning Meeting, October 1-2, 2007, Toronto, ON, Canada. “Data Integration, Analysis, and Visualization”*
8. Society for Neuro-oncology, November 16-18, 2007, Dallas, TX. “A TCGA GBM Data Primer”*
9. Biology of GBM TCGA Workshop, Mar 24-25, 2008, Bethesda, MD. “An update on GBM Analyses”*
10. 1st ICGC Scientific Planning Committee Meeting, November 15-17, 2008, Washington DC. “State of Ovarian Omics”*

11. AACR, April 21, 2009, Denver CO. NCI special session on TCGA. "A Preliminary Look at Data from the Serous Ovarian Carcinoma Project"*
12. Second ICGC Scientific Planning Committee Meeting, June 22-24, 2009, Hinxton, UK. "Update on TCGA"*
13. WNAR/IMS, June 14-17, 2009, Portland, OR. "Next-gen Sequencing and RNAseq Data: An Overview"*
14. Clinical Proteomic Technologies Strategy Workshop: Implementation of a New Cancer Biomarker Development Process, September 8-9, 2009, Bethesda, MD. "Subtypes and Mutations"*
15. Integrating Knowledge of Inherited Genetic Variation and Acquired Somatic Alterations in Cancer Tissues: Promising Scientific Opportunities, December 1-2, 2009, Bethesda, MD. "TCGA Ovarian Cancer Project"*
16. 2010 Workshop co-organizer on TCGA, at ISMB 2010
17. 3rd Scientific Workshop of the International Cancer Genome Consortium (ICGC), March 21-23, 2010, Madrid, Spain. "Large Scale Genomic Studies of Serous Ovarian Carcinoma"*
18. 4th Scientific Workshop of the International Cancer Genome Consortium (ICGC), December 1-3, 2010, Couran Cove, Australia, "The Ovarian Cancer Genome"*
19. AACR. Orlando FL April 3-6, 2011. "The effect of Technology Development on TCGA Direction"
20. San Antonio Breast Cancer Symposium, December 8-12, 2011, San Antonio, TX. "Impact of Systems Biology on Breast Cancer". Educational session
21. American Society for Clinical Oncology, June 1-5, 2012, Chicago, IL. "Impact of the Cancer Genome Atlas in Ovarian Cancer". Education session

Curriculum Development

1. OHSU Biostatistics Course Committee 2012

Development of Educational Resources

1. ICGC Bioinformatics and Data Sharing workgroups Toolbox development
Working with Dr. Lincoln Stein of OICR, supervise the development of an online reference for all topics relevant to ICGC participants, including but not limited to (1) white papers on best practices, (2) specific protocols, (3) curation of reference data sets (4) specification of standard definitions (5) specific special topics in sequencing, pathway analysis and data submission.

Mentoring and Advising

Postdoctoral and Predoctoral Mentoring:

I have completed mentoring thirteen post-doctoral fellows and five graduate students. I am currently training three doctoral students and three post-doctoral fellows.

1. Laura M. Heiser, Ph.D. Neuroscience. Postdoctoral fellow 2005-2008. Current position: Associate Professor, Oregon Health and Sciences University. Focus: development of cancer cell line signaling models that predict pathway activity.
2. Amrita Ray, Ph.D. Statistical Genetics. Postdoctoral fellow 2007-2009. Current

- position: Senior Data Scientist, Microsoft. Focus: development of methods to identify genetic influences on gene expression and copy number abnormalities in cancer.
3. Anguraj Sadananam, Ph.D. Cellular Biology. Postdoctoral fellow joint with Joe W. Gray 2007-2010. Current position: Team Leader ICR, Sutton, UK. Foci: (1) development of methods to identify fusion genes in breast cancer (2) identification of predictors of therapeutic response in cancer cell lines.
 4. Steffen Durinck, Ph.D. Bioengineering. Postdoctoral fellow 2008-2009. Current position: Scientist, Genentech. Focus: development of methods to analyze RNAseq data.
 5. William J. Gibb, Ph.D. Mechanical engineering. Postdoctoral fellow 2008-2010. Current position: Senior Bioinformatics Scientist, Genomic Health. Focus: Classification of breast cancer cell lines.
 6. Francois Pepin, Ph.D. Bioinformatics. Postdoctoral fellow 2010-2011. Current position: Bioinformatics Senior Manager, Roche Sequencing Solutions. Focus: development of methods to process shRNA synthetic lethality data.
 7. Anneleen Daemen, Ph.D. Computer Science. Postdoctoral fellow 2010-2011. Current position, Bioinformatics Scientist, Genentech. Focus: development of predictors of therapeutic response from cell line data.
 8. Obi L. Griffith, Ph.D. Bioinformatics. Postdoctoral fellow joint with Joe W. Gray 2010-2012. Current position, Assistant Professor, Washington University. Focus: Use of tumor profiling data to predict patient outcomes.
 9. Suzanne Fei, Ph.D. Bioinformatics. Postdoctoral fellow 2011-2015. Current position, Computational Genomics Core Director, OHSU. Focus: identification of genotypic influences on mutation spectra in kidney cancer.
 10. Amanda Esch, Ph.D. Post-doctoral fellow 2011-2014. Current position, Senior product application scientist, Fluidigm. Focus: development of mass tagged antibodies for multiplexed flow cytometry analyses of tumor cell signaling.
 11. Brett Johnson, Ph.D. Post-doctoral fellow 2015-2017. Focus: Genomic studies of pancreatic cancer. Current position staff at OHSU
 12. Ece Eksi, Ph.D. CEDAR post-doctoral fellow 2017-present. Focus: Single cells epigenomic analyses of prostate cancer.
 13. Josiah Wagner, Ph.D. CEDAR post-doctoral fellow. 2017-2021 Development of assays to detect cancer early. Current position staff at Providence Health.
 14. Kathryn Baker, PhD. CEDAR Post-doctoral fellow 2019-2021. Focus: use of ctDNA in cancer early detection. Current position Natera.
 15. Timothy Butler, graduate student, 2011-2016, post-doctoral fellow 2016-2017. Current position, postdoctoral fellow, Post-doctoral fellow Wellcome Trust Sanger Institute, Peter Campbell, mentor. Focus: Analysis of somatic mutations in rare tumors types; development of diagnostic applications of tumor burden/recurrence using unique DNA sequences.
 16. Asia Mitchell, graduate student, 2012-2018. Focus: Integration of large-scale data into networks to identify genomic causes of aberrant signaling in cancer. Current position Helix
 17. Michael Heskett, graduate student 2016-2021. Focus: Long non-coding RNAs that

regulate chromosome dynamics. Current position: Postdoctoral fellow with Christina Curtis at Stanford.

18. Burcu Gurun-Demir, graduate student 2019-2023. Focus: Selection of common drivers in clonal hematopoiesis. Current Position MD student at OHSU
 19. Chris Boniface 2019-2023. Focus: Use of ctDNA to study tumor biology. Current Position Postdoctoral Fellow OHSU
 20. Lora Iliev 2024- Focus: Genetics to Understand Disparities in Cancer
 21. Kevin Abuhanna 2024- Focus: Modeling Costs of Genetic Health Screening
 22. Toby Baker PhD 2024- Focus: Computational Modeling of Tumor Genome Evolution
 23. Katelyn Queen PhD- Focus: Statistical analysis of population genetics in the UCLA ATLAS program
 24. Maya Segal PhD- Focus: Statistical modeling for understanding tumor evolution
-

BIBLIOGRAPHY

RESEARCH PAPERS (PEER REVIEWED)

A. RESEARCH PAPERS – PEER REVIEWED

1. **Spellman PT**, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*. 9: 3273-97. 1998
2. Eisen MB, **Spellman PT**, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences, U.S.A.* 95: 14863-8. 1998 (I produced much of the data in the paper, discussed interpretation of the findings, and provided suggestions on displaying data).
3. Zhu G, **Spellman PT**, Volpe T, Brown PO, Botstein D, Davis TN, Futcher B. Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature*. 406: 90-4. 2000 (I produced gene expression profiling used in the paper)
4. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, **Spellman P**, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JC, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D, Brown PO. Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics*. 24: 227-35. 2000 (I created the website used in the paper)
5. Gasch AP, **Spellman PT**, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*. 11: 4241-57. 2000 (I generated a large fraction of the gene expression data used in the paper)
6. Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese J.C, Dwight SS, Kaloper M, Weng S, Jin H, Ball CA, Eisen MB, **Spellman PT** Brown PO, Botstein D, Cherry JM. The Stanford Microarray Database *Nucleic Acids Research*. 29: 152-5. 2001 (I helped build the initial implementations and feature sets for one of the first databases of large scale gene expression data).
7. George RA, Woolley JP, **Spellman PT**. Ceramic capillaries for use in microarray fabrication. *Genome Research*. 11: 1780-3. 2001
8. De Gregorio E, **Spellman PT**, Rubin GM, Lemaitre B. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proceedings of the National Academy of Sciences, U.S.A.* 98: 12590-5. 2001 (I generated a large fraction of the gene expression data used in the paper and led analyses of the gene expression data)
9. Brazma A, Hingamp P, Quackenbush J, Sherlock G, **Spellman P**, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics*. 29: 365-71. 2001 (I worked to identify key factors in gene expression data sharing)
10. **Spellman PT**, Rubin GM. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *Journal of Biology*. 1: 5. 2002
11. **Spellman PT**, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D,

- Sherlock G, Ball C, Lepage M, Swiatek M, Marks WL, Goncalves J, Markel S, Iordan D, Shojatalab M, Pizarro A, White J, Hubley R, Deutsch E, Senger M, Aronow BJ, Robinson A, Bassett D, Stoeckert CJ, Brazma A. Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biology* 3: RESEARCH. 0046. 2002
12. De Gregorio E, **Spellman PT**, Tzou P, Rubin GM, Lemaitre B. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *Embo Journal*. 21: 2568-79. 2002 (I generated a large fraction of the gene expression data used in the paper and led analyses of the gene expression data)
 13. Pile LA, **Spellman PT**, Katzenberger RJ, Wassarman DA. The SIN3 deacetylase complex represses genes encoding mitochondrial proteins: implications for the regulation of energy metabolism. *J Biol Chem*. 278: 37840-8. 2003 (I generated a large fraction of the gene expression data used in the paper and led analyses of the gene expression data)
 14. Vodovar N, Vinal M, Liehl P, Basset A, Degrouard J, **Spellman P**, Boccard F, Lemaitre B. *Drosophila* host defense after oral infection by an entomopathogenic *Pseudomonas* species. *PNAS*. 102:11414-9. 2005
 15. Brun S, Vidal S, **Spellman P**, Takahashi K, Tricoire H, Lemaitre B. The MAPKKK Mekk1 regulates the expression of Turandot stress genes in response to septic injury in *Drosophila*. *Genes Cells*. 11:397-407. 2006
 16. Jones AR, Pizarro A, **Spellman P**, Miller M. FuGE: Functional Genomics Object Model. *Omics*. 10:179-84. 2006 (I worked to design the data model)
 17. Chin K, DeVries KS, Fridlyand J, **Spellman PT**, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Daikree S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*. 6:529-41. 2006
 18. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, **Spellman PT**, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A, Gray JW. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 6:515-27. 2006
 19. Rayner TF, Rocca-Serra P, **Spellman PT**, Causton HC, Farne A, Holloway E, Irizarry RA, Liu J, Maier DS, Miller M, Petersen K, Quackenbush J, Sherlock G, Stoeckert CJ, White J, Whetzel PL, Wymore F, Parkinson H, Sarkans U, Ball CA, Brazma A. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics*. 7:489. 2006
 20. Jones AR, Miller M, Aebersold R, Apweiler R, Ball CA, Brazma A, Degreef J, Hardy N, Hermjakob H, Hubbard SJ, Hussey P, Igra M, Jenkins H, Julian RK, Laursen K, Oliver SG, Paton NW, Sansone SA, Sarkans U, Stoeckert CJ, Taylor CF, Whetzel PL, White JA, **Spellman P**, Pizarro A. The Functional Genomics Experiment model (FuGE): an extensible framework for standards in functional genomics. *Nat. Biotechnology*. 25:1127-33. 2007 (I worked to design the data model)
 21. Wang Y, Moorhead M, Karlin-Neumann G, Wang N, Ireland J, Lin S, Chen C, Heiser LM, Chin K, Esserman L, Gray JW, **Spellman PT**, Faham M. Performance of Molecular Inversion Probes (MIP) in Allele Copy Number Determination. *Genome Biology*. 8: R246. 2007 (I collaborated on interpreting the data)

22. Kenny PA, Lee GY, Myers R, Neve RM, Semeiks JR, **Spellman PT**, Lorenz K, Lee EH, Barcellos-Hoff MH, Petersen OW, Gray JW, Bissell MJ. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol Oncol*. 1:84-96. 2007
23. Brown LA, Kalloger SE, Miller MA, Shih IM, McKinney SE, Santos JL, Swenerton K, **Spellman PT**, Gray J, Gilks CB, Huntsman DG. Amplification of 11q13 in ovarian carcinoma. *Genes Chromosomes Cancer*. 47:481-9, 2008
24. Press JZ, De Luca A, Boyd N, Young S, Troussard A, Ridge Y, Kaurah P, Kalloger SE, Blood KA, Smith M, **Spellman PT**, Yang Y, Miller DM, Horsman D, Faham M, Gilks CB, Gray J, Huntsman DG. Ovarian carcinomas with genetic and epigenetic BRCA1 loss have distinct molecular abnormalities. *BMC Cancer*. 22:17. 2008
25. **The Cancer Genome Atlas Research Network**. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008
26. Mukherjee S, Pelech S, Neve RM, Kuo WL, Ziyad S, **Spellman PT**, Gray JW, Speed TP. Sparse combinatorial inference with an application in cancer biology. *Bioinformatics*. 2009
27. Durinck S, Bullard J, **Spellman PT**, Dudoit S. Integrated genomic visualization with R. BMC Bioinformatics. *GenomeGraphs*. 2009 (Dr Durinck was my trainee)
28. Hodgson JG, Yeh RF, Ray A, Wang NJ, Smirnov I, Yu M, Hariono S, Silber J, Feiler HS, Gray JW, **Spellman PT**, Vandenberg SR, Berger MS, James CD. Comparative analysis of gene copy number and mRNA expression in GBM tumors and GBM xenografts. *Neuro Oncology*. 2009
29. Mirzoeva OK, Das D, Heiser LM, Bhattacharya S, Siwak D, Gendelman R, Bayani N, Wang NJ, Neve RM, Guan Y, Hu Z, Knight Z, Feiler HS, Gascard P, Parvin B, **Spellman PT**, Shokat KM, Wyrobek AJ, Bissell MJ, McCormick F, Kuo WL, Mills GB, Gray JW, Korn WM. *Cancer Research*, 2009
30. Wang Y, Carlton VEH, Karlin-Neumann G, Sapolsky R, Zhang L, Moorhead M, Wang ZC, Richardson AL, Warren R, Walther A, Bondy M, Sahin A, Krahe R, Tuna M, Thompson PA, **Spellman PT**, Gray JW, Mills GB, Faham M. High quality copy number and genotype data from FFPE samples using Molecular Inversion Probe (MIP) microarrays. *BMC Medical Genomics*. 2009
31. Bengtsson H, Ray A, **Spellman P**, Speed TP. A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. *Bioinformatics*. 2009
32. Heiser LM, Wang NJ, Talcott CJ, Laderoute KR, Knapp M, Guan Y, Hu Z, Ziyad S, Weber BL, Laquerre S, Jackson JR, Wooster RF, Kuo WL, Gray JW, **Spellman PT**. *Genome Biology*. 2009
33. Durinck S, **Spellman PT**, Huber W. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. *Nature Protocols*. 2009 (Dr Durinck was my trainee)
34. Kuo WL, Das D, Ziyad S, Battacharya S, Gibb WJ, Heiser LM, Sadanandam A, Fontenay GV, Hu Z, Wang NJ, Bayani N, Feiler HS, Neve RM, Wyrobek AJ, **Spellman PT**, Marton LJ, Gray JW. A Systems Analysis of the Chemosensitivity of Breast Cancer Cells to the Polyamine Analogue PG-11047. *BMC Medicine*. 2009
35. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA,

- Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, **Spellman PT**, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN, Cancer Genome Atlas Network. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010
36. **International Cancer Genome Consortium**. International network of cancer genome projects. *Nature*. 2010
 37. Loss LA, Sadanandam A, Durinck S, Nautiya S, Flaucher D, Carlton VE, Moorhead M, Lu Y, Gray JW, Faham M, **Spellman P**, Parvin B. Prediction of epigenetically regulated genes in breast cancer cell lines. *BMC Bioinformatics*. 2010
 38. Lapuk A, Marr H, Jakkula L, Pedro H, Bhattacharya S, Purdom E, Hu Z, Simpson K, Pachter L, Wang N, Parvin B, Fontenay G, Speed T, Garbe J, Stampfer M, Bayandorian H, Dorton S, Clark TA, Schweitzer A, Wyrobek A, Feiler H, **Spellman P**, Conboy J, Gray JW. Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Mol Cancer Res*. 2010
 39. Nautiyal S, Carlton VE, Lu Y, Ireland JS, Moorhead M, Gray JW, **Spellman P**, Mindrinos M, Berg P, and Faham M. High-throughput method for analyzing methylation of CpGs in targeted genomic regions. *PNAS*. 2010
 40. Nabavizadeh N, Klifa C, Newitt, Lu DY, Chen YY, Hsu H, Fisher C, Tokayasu T, Olshen AB, **Spellman P**, Gray JW, Hylton N, Park CC. Topographic enhancement mapping of the cancer-associated breast stroma using breast MRI. *Integr. Biol*. 2011
 41. Wang XV, Verhaak RG, Purdom E, **Spellman PT**, Speed TP. Unifying gene expression measures from multiple platforms using factor analysis. *PLoS ONE*. 2011
 42. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, Feiler HE, Koh AH, Olshen AB, Danenberg KL, Tempero MA, **Spellman PT**, Hanahan D, Gray JW. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med*. 2011
 43. TCGA Network, **P.T. Spellman** corresponding author. Integrative Genomic Analyses of Ovarian Carcinoma. *Nature*. 2011
 44. Olshen AB, Bengtsson H, Neuvial P, **Spellman P**, Olshen RA, Seshan VE. Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. *Bioinformatics*. 2011
 45. Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, Pons J, Chan SW, Lam ET, Chu C, Park K, Hong SW, Hur JS, Huh N, Neuhaus IM, Yu SS, Grekin RT, Mauro TM, Cleaver JE, Kwok PY, LeBoit PE, Getz G, Cibulskis K, Aster JC, Huang H, Purdom E, Li J, Bolund L, Arron S, Gray JW, **Spellman PT***, Cho RJ* (*corresponding authors). Temporal Dissection of Tumorigenesis in Primary Cancers. *Cancer Discovery*. 2011
 46. Han J, Chang H, Loss LA, Zhang K, Baehner FL, Gray JW, **Spellman PT**, Parvin B. Comparison of sparse coding and kernel methods for histopathological classification of glioblastoma multiforme. *ISBI*. -714. 2011
 47. Chang H, Fontenay GV, Han J, Cong G, Baehner FL, Gray JW, **Spellman PT**, Parvin B. Morphometric Analysis of TCGA Glioblastoma Multiforme. *BMC Bioinformatics*. 2011 Dec 20; 12(1):484. [Epub ahead of print] PMID: 22185703.
 48. Wang NJ, Sanborn Z, Arnett KL, Bayston LJ, Liao W, Proby CM, Leigh IM, Collisson EA, Gordon PB, Jakkula L, Pennypacker S, Zou Y, Sharma M, North JP, Vemula SS, Mauro TM, Neuhaus IM, Leboit PE, Hur JS, Park K, Huh N, Kwok PY, Arron ST,

- Massion PP, Bale AE, Haussler D, Cleaver JE, Gray JW, **Spellman PT**, South AP, Aster JC, Blacklow SC, Cho RJ. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *PNAS*. 2011
49. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, Ng S, Gibb WJ, Wang NJ, Ziyad S, Tong F, Bayani N, Hu Z, Billig JJ, Dueregger A, Lewis S, Jakkula L, Korkola JE, Durinck S, Pepin F, Guan Y, Purdom E, Neuvial P, Bengtsson H, Wood KW, Smith PG, Vassilev LT, Hennessy BT, Greshock J, Bachman KE, Hardwicke MA, Park JW, Marton LJ, Wolf DM, Collisson EA, Neve RM, Mills GB, Speed TP, Feiler HS, Wooster RF, Haussler D, Stuart JM, Gray JW, **Spellman PT**. Subtype and pathway specific responses to anti-cancer compounds in breast cancer. *PNAS*. 2012
 50. Hill SM, Neve RM, Bayani N, Kuo WL, Ziyad S, **Spellman PT**, Gray JW, Mukherjee S. Integrating biological knowledge into variable selection: an empirical Bayes approach with an application to cancer biology. *BMC Bioinformatics*. 2012
 51. Creighton CJ, Hernandez-Herrera A, Jacobsen A, Levine DA, Mankoo P, Schultz N, Du Y, Zhang Y, Larsson E, Sheridan R, Xiao W, **Spellman PT**, Getz G, Wheeler DA, Perou CM, Gibbs RA, Sander C, Hayes DN, Gunaratne PH, The Cancer Genome Atlas Research Network. Integrated analyses of microRNAs demonstrate their widespread influence on gene expression in high-grade serous ovarian carcinoma. *PLoS One*. 2012
 52. Collisson EA, Trejo CL, Silva JM, Gu S, Korkola JE, Heiser LM, Charles RP, Rabinovich BA, Hann B, Dankort D, **Spellman P.T.**, Phillips WA, Gray JW, McMahon M. A Central Role for RAF ->MEK->ERK Signaling in the Genesis of Pancreatic Ductal Adenocarcinoma. *Cancer Discovery*. 2012
 53. **The Cancer Genome Atlas**. Comprehensive Molecular Characterization of Human Colon and Rectal Cancer. *Nature*. 2012
 54. Daemen A, Wolf DM, Korkola JE, Griffith OL, Frankum JR, Brough R, Jakkula LR, Wang NJ, Natrajan R, Reis-Filho JS, Lord CJ, Ashworth A, **Spellman PT**, Gray JW, Van't Veer LJ. Cross-platform pathway-based analysis identifies markers of response to the PARP inhibitor olaparib. *Breast Cancer Research Treatment*. 2012
 55. Hill SM, Lu Y, Molina J, Heiser LM, **Spellman PT**, Speed TP, Gray JW, Mills GB, Mukherjee S. Bayesian Inference of Signaling Network Topology in a Cancer Cell Line. *Bioinformatics*. 2012
 56. Patrillo LA, Wolf DM, Kapoun AM, Wang NJ, Barczak A, Xiao Y, Korkaya H, Baehner F, Lewicki J, Wicha M, Park JW, **Spellman PT**, Gray JW, Van't Veer L, Esserman LJ. Xenografts faithfully recapitulate breast cancer-specific gene expression patterns of parent primary breast tumors. *Breast Cancer Research Treatment*. 2012
 57. Chang H, Han J, **Spellman P**, Parvin B. Multi-Reference Level Set for Characterization of Nuclear Morphology in Glioblastoma Multiforme. *IEEE Transactions in Biomedical Engineering*. 2012
 58. **The Cancer Genome Atlas**. Comprehensive molecular portraits of human breast tumors. *Nature*. 2012
 59. Sadanandam A, Lal A, Benz SC, Eppenberger-Castori S, Scott G, Gray JW, **Spellman P**, Waldman F, Benz CC. Genomic aberrations in normal tissue adjacent to HER2-amplified breast cancers: field cancerization or contaminating tumor cells? *Breast Cancer Res Treat*. 2012
 60. Chang H, Han J, Borowsky A, Loss L, Gray JW, **Spellman PT**, Parvin B. Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and

- molecular association. *IEEE Trans Med Imaging*. 2013
61. Ordinario E, Han HJ, Furuta S, Heiser LM, Jakkula LR, Rodier F, **Spellman PT**, Campisi J, Gray JW, Bissell MJ, Kohwi Y, Kohwi-Shigematsu T. ATM suppresses SATB1-induced malignant progression in breast epithelial cells. *PLoS One*. 2012
 62. Verhaak RG, Tamayo P, Yang JY, Hubbard D, Zhang H, Creighton CJ, Fereday S, Lawrence M, Carter SL, Mermel CH, Kostic AD, Etemadmoghadam D, Saksena G, Cibulskis K, Duraisamy S, Levanon K, Sougnez C, Tsherniak A, Gomez S, Onofrio R, Gabriel S, Chin L, Zhang N, **Spellman PT**, Zhang Y, Akbani R, Hoadley KA, Kahn A, Köbel M, Huntsman D, Soslow RA, Defazio A, Birrer MJ, Gray JW, Weinstein JN, Bowtell DD, Drapkin R, Mesirov JP, Getz G, Levine D, Meyerson M. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma; Cancer Genome Atlas Research Network. *J Clin Invest*. 2013
 63. Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet*. 2013
 64. Kodigepalli KM, Anur P, **Spellman P**, Sims PJ, Nanjundan M. Phospholipid Scramblase 1, an interferon-regulated gene located at 3q23, is regulated by SnoN/SkiL in ovarian cancer cells. *Mol. Cancer* 2013.
 65. **TCGA Network**. Comprehensive molecular characterization of clear cell renal carcinoma. *Nature* 2013.
 66. Rhee JK, Kim K, Chae H, Evans J, Yan P, Zhang BT, Gray J, **Spellman P**, Huang TH, Nephew KH, Kim S. Integrated analysis of genome-wide DNA methylation and gene expression profiles in molecular subtypes of breast cancer. *NAR* 2013.
 67. Purdom E, Ho C, Grasso CS, Quist MJ, Cho RJ, **Spellman P**. Methods and challenges in timing chromosomal abnormalities within cancer samples. *Bioinformatics* 2013.
 68. **Cancer Genome Atlas Research Network**. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* 2013.
 69. Luoh SW, Ramsey B, Hanlon Newell A, Troxell M, Hu Z, Chin K, **Spellman P**, Olson S, Keenan E. HER-2 gene amplification in human breast cancer without concurrent HER-2 over-expression. Springerplus 2013.
 70. Daemen A, Griffith OL, Heiser LM, Wang NJ, Enache OM, Sanborn Z, Pepin F, Durinck S, Korkola JE, Griffith M, Hur JS, Huh N, Chung J, Cope L, Fackler MJ, Umbricht C, Sukumar S, Seth P, Sukhatme VP, Jakkula LR, Lu Y, Mills GB, Cho RJ, Collisson EA, Van't Veer LJ, **Spellman PT**, Gray JW. Modeling precision treatment of breast cancer. *Genome Biology* 2013.
 71. Brennan CW, Verhaak RG, McKenna A, et al. and TCGA Research Network. The somatic genomic landscape of glioblastoma. *Cell*. 2013.
 72. Griffith OL, Pepin F, Enache OM, Heiser LM, Collisson EA, **Spellman PT**, Gray JW. A robust prognostic signature for hormone-positive node-negative breast cancer. *Genome Medicine* 2013.
 73. Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, Zhang Q, Koboldt DC, Xie M, Kandoth C, McMichael JF, Wyczalkowski MA, Larson DE, Schmidt HK, Miller CA, Fulton RS, **Spellman PT**, Mardis ER, Druley TE, Graubert TA, Goodfellow PJ, Raphael BJ, Wilson RK, Ding L. Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun*. 2014.
 74. Nayak N, Chang H, Borowsky A, **Spellman P**, Parvin B. CLASSIFICATION OF TUMOR HISTOPATHOLOGY VIA SPARSE FEATURE LEARNING. *Proc IEEE Int*

- Symp Biomed Imaging. 2013 Apr;2013.
75. Chang H, Borowsky A, **Spellman P**, Parvin B. Classification of Tumor Histology via Morphometric Context. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2013
 76. Chang H, Nayak N, **Spellman PT**, Parvin B. Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching. *Med Image Comput Comput Assist Interv*. 2013
 77. Boutros PC, Ewing AD, Ellrott K, Norman TC, Dang KK, Hu Y, Kellen MR, Suver C, Bare JC, Stein LD, **Spellman PT**, Stolovitzky G, Friend SH, Margolin AA, Stuart JM. Global optimization of somatic variant identification in cancer genomes with a global community challenge. *Nat Genet*. 2014
 78. Fu X, Creighton CJ, Biswal NJ, Kumar V, Shea M, Herrera S, Contreras A, Gutierrez C, Wang T, Nanda S, Giuliano M, Morrison G, Nardone A, Karlin KL, Westbrook TF, Heiser LM, Anur P, **Spellman P**, Guichard SM, Smith PD, Davies BR, Klinowska T, Lee AV, Mills GB, Rimawi MF, Hilsenbeck SG, Gray JW, Joshi A, Osborne C, Schiff R. Overcoming endocrine resistance due to reduced PTEN levels in estrogen receptor-positive breast cancer by co-targeting mammalian target of rapamycin, protein kinase B, or mitogen-activated protein kinase kinase.
 79. Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, et al.. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell*. 2014 Sep 8
 80. Hoadley KA, Yau C, Wolf DM, Cherniack AD, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014 Aug 14;158(4):929-44. doi: 10.1016/j.cell.2014.06.049. Epub 2014 Aug 7.
 81. Zhou Y, Chang H, Barner K, **Spellman P**, Parvin B. Classification of Histology Sections via Multispectral Convolutional Sparse Coding Conf Comput Vis Pattern Recognit Workshops. 2014 Jun.
 82. Zheng CL, Wang NJ, Chung J, Moslehi H, Sanborn JZ, Hur JS, Collisson EA, Vemula SS, Naujokas A, Chiotti KE, Cheng JB, Fassihi H, Blumberg AJ, Bailey CV, Fudem GM, Mihm FG, Cunningham BB, Neuhaus IM, Liao W, Oh DH, Cleaver JE, LeBoit PE, Costello JF, Lehmann AR, Gray JW, **Spellman PT**, Arron ST, Huh N, Purdom E, Cho RJ. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Rep*. 2014
 83. Rozanov D, **Spellman P**, Savinov A, Strongin AY. A Humanized Leucine Zipper-TRAIL Hybrid Induces Apoptosis of Tumors both In Vitro and In Vivo. *PLoS One*. 2015 Apr 7.
 84. Grasso CS, Tang Y, Truffaux N, Berlow NE, et al. Functionally defined therapeutic targets in diffuse intrinsic pontine glioma. *Nature Medicine*. 2015 May 4.
 85. Butler TM, Johnson-Camacho K, Peto M, Wang NJ, Macey TA, Korkola JE, Koppie TM, Corless CL, Gray JW, **Spellman PT**. Exome Sequencing of Cell-Free DNA from Metastatic Cancer Patients Identifies Clinically Actionable Mutations Distinct from Primary Disease. *PLoS One*. 2015;
 86. Sanborn JZ, Chung J, Purdom E, Wang NJ, Kakavand H, Wilmott JS, Butler T, Thompson JF, et al. Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination. *Proc Natl Acad Sci U S A*. 2015 September 1
 87. Korkola JE, Collisson EA, Heiser L, Oates C, Bayani N, Itani S, Esch A, Thompson W, Griffith OL, Wang NJ, Kuo WL, Cooper B, Billig J, Ziyad S, Hung JL, Jakkula L, Feiler H, Lu Y, Mills GB, **Spellman PT**, Tomlin C, Mukherjee S, Gray JW. Decoupling of the

- PI 3K Pathway via Mutation Necessitates Combinatorial Treatment in *HER2+ Breast Cancer*. PLoS One. 2015; 10(7)
88. Linehan WM*, **Spellman PT***, Ricketts CJ, Creighton CJ, Fei SS, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N Engl J Med. 2015 Nov 4
 89. Liu Y, Yasukawa M, Chen K, Hu L, Broaddus RR, Ding L, Mardis ER, **Spellman P**, Levine DA, Mills GB, Shmulevich I, Sood AK, Zhang W. Association of Somatic Mutations of ADAMTS Genes With Chemotherapy Sensitivity and Survival in High-Grade Serous Ovarian Carcinoma. JAMA Oncol. 2015 Jul
 90. Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, Zhang Y, Sokolov A, Paull EO, Wong CK, Graim K, Bivol A, Wang H, Zhu F, Afsari B, Danilova LV, Favorov AV, Lee WS, Taylor D, Hu CW, Long BL, Noren DP, Bisberg AJ; HPN-DREAM Consortium, Mills GB, Gray JW, Kellen M, Norman T, Friend S, Qutub AA, Fertig EJ, Guan Y, Song M, Stuart JM, **Spellman PT**, Koepl H, Stolovitzky G, Saez-Rodriguez J, Mukherjee S. Inferring causal molecular networks: empirical assessment through a community-based effort. Nat Methods. 2016 Feb 22. doi: 10.1038/nmeth.3773. [Epub ahead of print]
 91. Fleyshman D, Cheney P, Ströse A, Mudambi S, Safina A, Commane M, Purmal A, Morgan K, Wang NJ, Gray J, **Spellman PT**, Issaeva N, Gurova K. ARTIK-52 induces replication-dependent DNA damage and p53 activation exclusively in cells of prostate and breast cancer origin. Cell Cycle. 2016 Feb;15(3):455-70. doi: 10.1080/15384101.2015.1127478. Epub 2015 Dec 22. PMID: 26694952
 92. Ivanov M, Baranova A, Butler T, **Spellman P**, Mileyko V. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. BMC Genomics. 2015 Dec 16;16 Suppl 13:S1. doi: 10.1186/1471-2164-16-S13-S1. Epub 2015 Dec 16. PMID: 26693644
 93. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, Heisler LE, Beck TA, Simpson JT, Tonon L, Sertier AS, Patch AM, Jäger N, Ginsbach P, Drews R, Paramasivam N, Kabbe R, Chotewutmontri S, Diessl N, Previti C, Schmidt S, Brors B, Feuerbach L, Heinold M, Gröbner S, Korshunov A, Tarpey PS, Butler AP, Hinton J, Jones D, Menzies A, Raine K, Shepherd R, Stebbings L, Teague JW, Ribeca P, Giner FC, Beltran S, Raineri E, Dabad M, Heath SC, Gut M, Denroche RE, Harding NJ, Yamaguchi TN, Fujimoto A, Nakagawa H, Quesada V, Valdés-Mas R, Nakken S, Vodák D, Bower L, Lynch AG, Anderson CL, Waddell N, Pearson JV, Grimmond SM, Peto M, **Spellman P**, He M, Kandoth C, Lee S, Zhang J, Létourneau L, Ma S, Seth S, Torrents D, Xi L, Wheeler DA, López-Otín C, Campo E, Campbell PJ, Boutros PC, Puente XS, Gerhard DS, Pfister SM, McPherson JD, Hudson TJ, Schlesner M, Lichter P, Eils R, Jones DT, Gut IG. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. Nat Commun. 2015 Dec 9;6:10001. doi: 10.1038/ncomms10001.
 94. Lue HW, Cole B, Rao SA, Podalak J, Van Gaest A, King C, Eide CA, Wilmot B, Xue C, **Spellman PT**, Heiser LM, Tyner JW, Thomas GV. Src and STAT3 inhibitors synergize to promote tumor inhibition in renal cell carcinoma. Oncotarget. 2015 Nov 26. doi: 10.18632/oncotarget.5971.
 95. Fleyshman D, Cheney P, Ströse A, Mudambi S, Safina A, Commane M, Purmal A, Morgan K, Wang NJ, Gray J, **Spellman PT**, Issaeva N, Gurova K. ARTIK-52 induces

- replication-dependent DNA damage and p53 activation exclusively in cells of prostate and breast cancer origin. *Cell Cycle*. 2016 Feb;15(3):455-70. doi: 10.1080/15384101.2015.1127478. Epub 2015 Dec 22.
96. Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, Zhang Y, Sokolov A, Paull EO, Wong CK, Graim K, Bivol A, Wang H, Zhu F, Afsari B, Danilova LV, Favorov AV, Lee WS, Taylor D, Hu CW, Long BL, Noren DP, Bisberg AJ; HPN-DREAM Consortium, Mills GB, Gray JW, Kellen M, Norman T, Friend S, Qutub AA, Fertig EJ, Guan Y, Song M, Stuart JM, **Spellman PT**, Koeppl H, Stolovitzky G, Saez-Rodriguez J, Mukherjee S. Inferring causal molecular networks: empirical assessment through a community-based effort. *Nat Methods*. 2016 Apr;13(4):310-8. doi: 10.1038/nmeth.3773. Epub 2016 Feb 22.
 97. Fei SS, Mitchell AD, Heskett MB, Vocke CD, Ricketts CJ, Peto M, Wang NJ, Sönmez K, Linehan WM, **Spellman PT**. Patient-specific factors influence somatic variation patterns in von Hippel-Lindau disease renal tumours. *Nat Commun*. 2016 May 13;7:11588. doi: 10.1038/ncomms11588.
 98. Munson DJ, Egelston CA, Chiotti KE, Parra ZE, Bruno TC, Moore BL, Nakano TA, Simons DL, Jimenez G, Yim JH, Rozanov DV, Falta MT, Fontenot AP, Reynolds PR, SM, Borges VF, Kappler JW, **Spellman PT**, Lee PP, Slansky JE. Identification of shared TCR sequences from T cells in human breast cancer using emulsion RT-PCR. *Proc Natl Acad Sci U S A*. 2016 Jul 19;113(29):8272-7. doi:10.1073/pnas.1606994113. Epub 2016 Jun 15. PubMed PMID: 27307436; PubMed Central PMCID: PMC4961128.
 99. Hu Z, Mao JH, Curtis C, Huang G, Gu S, Heiser L, Lenburg ME, Korkola JE, Bayani N, Samarajiwa S, Seoane JA, Dane MA, Esch A, Feiler HS, Wang NJ, Hardwicke MA, Laquerre S, Jackson J, W Wood K, Weber B, **Spellman PT**, Aparicio S, Wooster R, Caldas C, Gray JW. Genome co-amplification upregulates a mitotic gene network activity that predicts outcome and response to mitotic protein inhibitors in breast cancer. *Breast Cancer Res*. 2016 Jul 1;18(1):70. doi:10.1186/s13058-016-0728-y. Erratum in: *Breast Cancer Res*. 2017 Feb 9;19(1):17. PubMed PMID: 27368372; PubMed Central PMCID: PMC4930593.
 100. Fu X, Jeselsohn R, Pereira R, Hollingsworth EF, Creighton CJ, Li F, Shea M, Nardone A, De Angelis C, Heiser LM, Anur P, Wang N, Grasso CS, **Spellman PT**, Griffith OL, Tsimelzon A, Gutierrez C, Huang S, Edwards DP, Trivedi MV, Rimawi MF, Lopez-Terrada D, Hilsenbeck SG, Gray JW, Brown M, Osborne CK, Schiff R. FOXA1 overexpression mediates endocrine resistance by altering the ER transcriptome and IL-8 expression in ER-positive breast cancer. *Proc Natl Acad Sci U S A*. 2016 Oct 25;113(43):E6600-E6609. Epub 2016 Oct 6. PubMed PMID: 27791031; PubMed Central PMCID: PMC5087040.
 101. Van Hook K, Wang Z, Chen D, Nold C, Zhu Z, Anur P, Lee HJ, Yu Z, Sheppard B, Dai MS, Sears R, **Spellman P**, Lopez CD. Δ N-ASPP2, a novel isoform of the ASPP2 tumor suppressor, promotes cellular survival. *Biochem Biophys Res Commun*. 2017 Jan 22;482(4):1271-1277. doi: 10.1016/j.bbrc.2016.12.027. Epub 2016 Dec 8. PubMed PMID: 27939881; PubMed Central PMCID: PMC5237613.
 102. Hill SM, Nesser NK, Johnson-Camacho K, Jeffress M, Johnson A, Boniface C, Spencer SE, Lu Y, Heiser LM, Lawrence Y, Pande NT, Korkola JE, Gray JW, Mills GB, Mukherjee S, **Spellman PT**. Context Specificity in Causal Signaling Networks Revealed by Phosphoprotein Profiling. *Cell Syst*. 2017 Jan 25;4(1):73-83.e10. doi:

- 10.1016/j.cels.2016.11.013. Epub 2016 Dec 22. PubMed PMID: 28017544; PubMed Central PMCID: PMC5279869.
103. Guo Q, Minnier J, Burchard J, Chiotti K, **Spellman P**, Schedin P. Physiologically activated mammary fibroblasts promote postpartum mammary cancer. *JCI Insight*. 2017 Mar 23;2(6):e89206. doi: 10.1172/jci.insight.89206. PubMed PMID: 28352652; PubMed Central PMCID: PMC5358521.
 104. **Cancer Genome Atlas Research Network**. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell*. 2017 Aug 14;32(2):185-203.e13. doi: 10.1016/j.ccell.2017.07.007. PubMed PMID: 28810144.
 105. Korkola JE, Collisson EA, Heiser LM, Oates C, Bayani N, Itani S, Esch A, Thompson W, Griffith OL, Wang NJ, Kuo WL, Cooper B, Billig J, Ziyad S, Hung JL, Jakkula L, Feiler H, Lu Y, Mills GB, **Spellman PT**, Tomlin C, Mukherjee S, Gray JW. Correction: Decoupling of the PI3K Pathway via Mutation Necessitates Combinatorial Treatment in HER2+ Breast Cancer. *PLoS One*. 2017 Oct 11;12(10):e0186551. doi: 10.1371/journal.pone.0186551. eCollection 2017. PubMed PMID: 29020035.
 106. Wang Y, Dorrell C, Naugler WE, Heskett M, **Spellman P**, Li B, Galivo F, Haft A, Wakefield L, Grompe M. Long-Term Correction of Diabetes in Mice by In Vivo Reprogramming of Pancreatic Ducts. *Mol Ther*. 2018 Feb 21. pii: S1525-0016(18)30067-4. doi: 10.1016/j.ymthe.2018.02.014. [Epub ahead of print] PMID:29550076
 107. The Integrated Genomic Landscape of Thymic Epithelial Tumors. Radovich M, Pickering CR, Felau I, Ha G, Zhang H, Jo H, Hoadley KA, Anur P, Zhang J, McLellan M, Bowlby R, Matthew T, Danilova L, Hegde AM, Kim J, Leiserson MDM, Sethi G, Lu C, Ryan M, Su X, Cherniack AD, Robertson G, Akbani R, **Spellman P**, Weinstein JN, Hayes DN, Raphael B, Lichtenberg T, Leraas K, Zenklusen JC; Cancer Genome Atlas Network, Fujimoto J, Scapulatempo-Neto C, Moreira AL, Hwang D, Huang J, Marino M, Korst R, Giaccone G, Gokmen-Polar Y, Badve S, Rajan A, Ströbel P, Girard N, Tsao MS, Marx A, Tsao AS, Loehrer PJ. *Cancer Cell*. 2018 Feb 12;33(2):244-258.e10. doi: 10.1016/j.ccell.2018.01.003. PMID:29438696
 108. Rozanov DV, Rozanov ND, Chiotti KE, Reddy A, Wilmarth PA, David LL, Cha SW, Woo S, Pevzner P, Bafna V, Burrows GG, Rantala JK, Levin T, Anur P, Johnson-Camacho K, Tabatabaei S, Munson DJ, Bruno TC, Slansky JE, Kappler JW, Hirano N, Boegel S, Fox BA, Egelston C, Simons DL, Jimenez G, Lee PP, Gray JW, **Spellman PT**. MHC class I loaded ligands from breast cancer cell lines: A potential HLA-I-typed antigen collection. *J Proteomics*. 2018 Mar 30;176:13-23. doi: 10.1016/j.jprot.2018.01.004. Epub 2018 Jan 10. PMID:29331515
 109. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. **Cancer Genome Atlas Research Network**. *Cell*. 2017 Nov 2;171(4):950-965.e28. doi: 10.1016/j.cell.2017.10.014. PMID:29100075
 110. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H,

- Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CE; **Cancer Genome Atlas Research Network**, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG, Shmulevich L. *The Immune Landscape of Cancer*. *Immunity*. 2018 Apr 4. pii: S1074-7613(18)30121-3. doi: 10.1016/j.immuni.2018.03.023. [Epub ahead of print] PubMed PMID: 29628290.
111. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD, Thorsson V; **Cancer Genome Atlas Research Network**, Hu H. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell*. 2018 Apr 5;173(2):400-416.e11. doi: 10.1016/j.cell.2018.02.052. PubMed PMID: 29625055.
 112. Chen H, Li C, Peng X, Zhou Z, Weinstein JN; **Cancer Genome Atlas Research Network**, Liang H. A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell*. 2018 Apr 5;173(2):386-399.e12. doi:10.1016/j.cell.2018.03.027. PubMed PMID: 29625054; PubMed Central PMCID:PMC5890960.
 113. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, Ng PK, Jeong KJ, Cao S, Wang Z, Gao J, Gao Q, Wang F, Liu EM, Mularoni L, Rubio-Perez C, Nagarajan N, Cortés-Ciriano I, Zhou DC, Liang WW, Hess JM, Yellapantula VD, Tamborero D, Gonzalez-Perez A, Suphavitai C, Ko JY, Khurana E, Park PJ, Van Allen EM, Liang H; MC3 Working Group; **Cancer Genome Atlas Research Network**, Lawrence MS, Godzik A, Lopez-Bigas N, Stuart J, Wheeler D, Getz G, Chen K, Lazar AJ, Mills GB, Karchin R, Ding L. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*. 2018 Apr 5;173(2):371-385.e18. doi: 10.1016/j.cell.2018.02.060. PubMed PMID: 29625053.
 114. Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, Paczkowska M, Reynolds S, Wyczalkowski MA, Oak N, Scott AD, Krassowski M, Cherniack AD, Houlihan KE, Jayasinghe R, Wang LB, Zhou DC, Liu D, Cao S, Kim YW, Koiré A, McMichael JF, Huchtagowder V, Kim TB, Hahn A, Wang C, McLellan MD, Al-Mulla F, Johnson KJ; **Cancer Genome Atlas Research Network**, Lichtarge O, Boutros PC, Raphael B, Lazar AJ, Zhang W, Wendl MC, Govindan R, Jain S, Wheeler D, Kulkarni S, Dipersio JF, Reimand J, Meric-Bernstam F, Chen K, Shmulevich I, Plon SE, Chen F, Ding L. Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell*. 2018 Apr 5;173(2):355-370.e14. doi: 10.1016/j.cell.2018.03.039. PubMed PMID: 29625052.
 115. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwińska P, Mazurek S, Mishra L, Heyn H, Krasnitz A, Godwin AK, Lazar AJ; **Cancer Genome Atlas Research Network**, Stuart JM, Hoadley KA, Laird PW, Noushmehr H, Wiznerowicz M. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell*. 2018 Apr 5;173(2):338-354.e15. doi:10.1016/j.cell.2018.03.034. PubMed PMID: 29625051.
 116. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadou S, Liu DL, Kantheti HS, Saghafeina S, Chakravarty D, Daian F, Gao Q, Bailey MH, Liang WW, Foltz SM, Shmulevich I, Ding L, Heins Z, Ochoa A, Gross B, Gao J, Zhang H, Kundra R, Kandoth C, Bahceci I, Dervishi L, Dogrusoz U, Zhou W, Shen H, Laird PW, Way GP, Greene CS, Liang H, Xiao Y, Wang C, Iavarone A, Berger AH, Bivona TG, Lazar AJ, Hammer GD, Giordano T, Kwong LN, McArthur G, Huang C, Tward AD, Frederick MJ, McCormick F, Meyerson M; **Cancer Genome Atlas Research**

- Network**, Van Allen EM, Cherniack AD, Ciriello G, Sander C, Schultz N. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell*. 2018 Apr 5;173(2):321-337.e10. doi:10.1016/j.cell.2018.03.035. PubMed PMID: 29625050.
117. Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, Gibbs DL, Weerasinghe A, Huang KL, Tokheim C, Cortés-Ciriano I, Jayasinghe R, Chen F, Yu L, Sun S, Olsen C, Kim J, Taylor AM, Cherniack AD, Akbani R, Suphavitai C, Nagarajan N, Stuart JM, Mills GB, Wyczalkowski MA, Vincent BG, Hutter CM, Zenklusen JC, Hoadley KA, Wendl MC, Shmulevich L, Lazar AJ, Wheeler DA, Getz G; **Cancer Genome Atlas Research Network**. Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell*. 2018 Apr 5;173(2):305-320.e10. doi: 10.1016/j.cell.2018.03.033. PubMed PMID: 29625049.
 118. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V, Akbani R, Bowlby R, Wong CK, Wiznerowicz M, Sanchez-Vega F, Robertson AG, Schneider BG, Lawrence MS, Noushmehr H, Malta TM; **Cancer Genome Atlas Network**, Stuart JM, Benz CC, Laird PW. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. 2018 Apr 5;173(2):291-304.e6. doi: 10.1016/j.cell.2018.03.022. PubMed PMID: 29625048.
 119. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, Seoane JA, Farshidfar F, Bowlby R, Islam M, Kim J, Chatila W, Akbani R, Kanchi RS, Rabkin CS, Willis JE, Wang KK, McCall SJ, Mishra L, Ojesina AI, Bullman S, Pedamallu CS, Lazar AJ, Sakai R; **Cancer Genome Atlas Research Network**, Thorsson V, Bass AJ, Laird PW. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell*. 2018 Apr 9;33(4):721-735.e8. doi: 10.1016/j.ccell.2018.03.010. Epub 2018 Apr 2. PubMed PMID: 29622466.
 120. Wang Z, Yang B, Zhang M, Guo W, Wu Z, Wang Y, Jia L, Li S; **Cancer Genome Atlas Research Network**, Xie W, Yang D. lncRNA Epigenetic Landscape Analysis Identifies EPIC1 as an Oncogenic lncRNA that Interacts with MYC and Promotes Cell-Cycle Progression in Cancer. *Cancer Cell*. 2018 Apr 9;33(4):706-720.e9. doi: 10.1016/j.ccell.2018.03.006. Epub 2018 Apr 2. PubMed PMID: 29622465.
 121. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, Liu Y, Fan H, Shen H, Ravikumar V, Rao A, Schultz A, Li X, Sumazin P, Williams C, Mestdagh P, Gunaratne PH, Yau C, Bowlby R, Robertson AG, Tiezzi DG, Wang C, Cherniack AD, Godwin AK, Kuderer NM, Rader JS, Zuna RE, Sood AK, Lazar AJ, Ojesina AI, Adebamowo C, Adebamowo SN, Baggerly KA, Chen TW, Chiu HS, Lefever S, Liu L, MacKenzie K, Orsulic S, Roszik J, Shelley CS, Song Q, Vellano CP, Wentzensen N; **Cancer Genome Atlas Research Network**, Weinstein JN, Mills GB, Levine DA, Akbani R. A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell*. 2018 Apr 9;33(4):690-705.e9. doi: 10.1016/j.ccell.2018.03.014. Epub 2018 Apr 2. PubMed PMID: 29622464.
 122. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J, Lazar AJ; **Cancer Genome Atlas Research Network**, Cherniack AD, Beroukhi R, Meyerson M. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell*. 2018 Apr 9;33(4):676-689.e3. doi: 10.1016/j.ccell.2018.03.007. Epub 2018 Apr 2. PubMed PMID: 29622463.
 123. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, Bowlby R, Gibb

- EA, Akbani R, Beroukhir R, Bottaro DP, Choueiri TK, Gibbs RA, Godwin AK, Haake S, Hakimi AA, Henske EP, Hsieh JJ, Ho TH, Kanchi RS, Krishnan B, Kwaitkowski DJ, Lui W, Merino MJ, Mills GB, Myers J, Nickerson ML, Reuter VE, Schmidt LS, Shelley CS, Shen H, Shuch B, Signoretti S, Srinivasan R, Tamboli P, Thomas G, Vincent BG, Vocke CD, Wheeler DA, Yang L, Kim WT, Robertson AG; Cancer Genome Atlas Research Network, **Spellman PT**, Rathmell WK, Linehan WM. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep*. 2018 Apr 3;23(1):313-326.e5. doi: 10.1016/j.celrep.2018.03.075. PubMed PMID: 29617669.
124. Chiu HS, Somvanshi S, Patel E, Chen TW, Singh VP, Zorman B, Patil SL, Pan Y, Chatterjee SS; **Cancer Genome Atlas Research Network**, Sood AK, Gunaratne PH, Sumazin P. Pan-Cancer Analysis of lnc RNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. *Cell Rep*. 2018 Apr 3;23(1):297-312.e12. doi: 10.1016/j.celrep.2018.03.064. PubMed PMID: 29617668.
 125. Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG; **Cancer Genome Atlas Research Network**, Buonamici S, Yu L. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep*. 2018 Apr 3;23(1):282-296.e4. doi: 10.1016/j.celrep.2018.01.088. PubMed PMID: 29617667.
 126. Jayasinghe RG, Cao S, Gao Q, Wendl MC, Vo NS, Reynolds SM, Zhao Y, Climente-González H, Chai S, Wang F, Varghese R, Huang M, Liang WW, Wyczalkowski MA, Sengupta S, Li Z, Payne SH, Fenyö D, Miner JH, Walter MJ; **Cancer Genome Atlas Research Network**, Vincent B, Eyraas E, Chen K, Shmulevich I, Chen F, Ding L. Systematic Analysis of Splice-Site-Creating Mutations in Cancer. *Cell Rep*. 2018 Apr 3;23(1):270-281.e3. doi: 10.1016/j.celrep.2018.03.052. PubMed PMID: 29617666.
 127. Peng X, Chen Z, Farshidfar F, Xu X, Lorenzi PL, Wang Y, Cheng F, Tan L, Mojumdar K, Du D, Ge Z, Li J, Thomas GV, Birsoy K, Liu L, Zhang H, Zhao Z, Marchand C, Weinstein JN; **Cancer Genome Atlas Research Network**, Bathe OF, Liang H. Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers. *Cell Rep*. 2018 Apr 3;23(1):255-269.e4. doi: 10.1016/j.celrep.2018.03.077. PubMed PMID: 29617665.
 128. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, Fan H, Shen H, Way GP, Greene CS, Liu Y, Akbani R, Feng B, Donehower LA, Miller C, Shen Y, Karimi M, Chen H, Kim P, Jia P, Shinbrot E, Zhang S, Liu J, Hu H, Bailey MH, Yau C, Wolf D, Zhao Z, Weinstein JN, Li L, Ding L, Mills GB, Laird PW, Wheeler DA, Shmulevich I; **Cancer Genome Atlas Research Network**, Monnat RJ Jr., Xiao Y, Wang C. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across **The Cancer Genome Atlas**. *Cell Rep*. 2018 Apr 3;23(1):239-254.e6. doi:10.1016/j.celrep.2018.03.076. PubMed PMID: 29617664.
 129. Gao Q, Liang WW, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, Liao WW, Reynolds SM, Wyczalkowski MA, Yao L, Yu L, Sun SQ; Fusion Analysis Working Group; **Cancer Genome Atlas Research Network**, Chen K, Lazar AJ, Fields RC, Wendl MC, Van Tine BA, Vij R, Chen F, Nykter M, Shmulevich I, Ding L. Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. *Cell Rep*. 2018 Apr 3;23(1):227-238.e3. doi: 10.1016/j.celrep.2018.03.050. PubMed

PMID: 29617662.

130. Ge Z, Leighton JS, Wang Y, Peng X, Chen Z, Chen H, Sun Y, Yao F, Li J, Zhang H, Liu J, Shriver CD, Hu H; **Cancer Genome Atlas Research Network**, Piwnicka-Worms H, Ma L, Liang H. Integrated Genomic Analysis of the Ubiquitin Pathway across Cancer Types. *Cell Rep.* 2018 Apr 3;23(1):213-226.e3. doi:10.1016/j.celrep.2018.03.047. PubMed PMID: 29617661.
131. Campbell JD, Yau C, Bowlby R, Liu Y, Brennan K, Fan H, Taylor AM, Wang C, Walter V, Akbani R, Byers LA, Creighton CJ, Coarfa C, Shih J, Cherniack AD, Gevaert O, Prunello M, Shen H, Anur P, Chen J, Cheng H, Hayes DN, Bullman S, Pedamallu CS, Ojesina AI, Sadeghi S, Mungall KL, Robertson AG, Benz C, Schultz A, Kanchi RS, Gay CM, Hegde A, Diao L, Wang J, Ma W, Sumazin P, Chiu HS, Chen TW, Gunaratne P, Donehower L, Rader JS, Zuna R, Al-Ahmadie H, Lazar AJ, Flores ER, Tsai KY, Zhou JH, Rustgi AK, Drill E, Shen R, Wong CK; **Cancer Genome Atlas Research Network**, Stuart JM, Laird PW, Hoadley KA, Weinstein JN, Peto M, Pickering CR, Chen Z, Van Waes C. Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. *Cell Rep.* 2018 Apr 3;23(1):194-212.e6. doi: 10.1016/j.celrep.2018.03.063. PubMed PMID: 29617660.
132. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, Samaras D, Shroyer KR, Zhao T, Batiste R, Van Arnam J; **Cancer Genome Atlas Research Network**, Shmulevich I, Rao AUK, Lazar AJ, Sharma A, Thorsson V. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Rep.* 2018 Apr 3;23(1):181-193.e7. doi:10.1016/j.celrep.2018.03.086. PubMed PMID: 29617659.
133. Way GP, Sanchez-Vega F, La K, Armenia J, Chatila WK, Luna A, Sander C, Cherniack AD, Mina M, Ciriello G, Schultz N; **Cancer Genome Atlas Research Network**, Sanchez Y, Greene CS. Machine Learning Detects Pan-cancer Ras Pathway Activation in The Cancer Genome Atlas. *Cell Rep.* 2018 Apr 3;23(1):172-180.e3. doi: 10.1016/j.celrep.2018.03.046. PubMed PMID: 29617658.
134. Schaub FX, Dhankani V, Berger AC, Trivedi M, Richardson AB, Shaw R, Zhao W, Zhang X, Ventura A, Liu Y, Ayer DE, Hurlin PJ, Cherniack AD, Eisenman RN, Bernard B, Grandori C; **Cancer Genome Atlas Network**. Pan-cancer Alterations of the MYC Oncogene and Its Proximal Network across the Cancer Genome Atlas. *Cell Syst.* 2018 Mar 28;6(3):282-300.e2. doi: 10.1016/j.cels.2018.03.003. PubMed PMID: 29596783; PubMed Central PMCID: PMC5892207.
135. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, Hess J, Ma S, Chiotti KE, McLellan M, Sofia HJ, Hutter C, Getz G, Wheeler D, Ding L; MC3 Working Group; **Cancer Genome Atlas Research Network**. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Syst.* 2018 Mar 28;6(3):271-281.e7. doi: 10.1016/j.cels.2018.03.002. PubMed PMID: 29596782.
136. Tarabichi M, Martincorena I, Gerstung M, Leroi AM, The PCAWG Evolution and Heterogeneity Working Group, **Spellman PT**, Morris QD, Lingjærde OC, Wedge DC, Van Loo P. Neutral Tumor Evolution. *Nature Genetics.* 2018 Oct 29; doi: 10.1038/s41588-018-0258-x
137. Wang Y, Xu X, Maglic D, Dill MT, Mojumdar K, Ng PK, Jeong KJ, Tsang YH, Moreno D, Bhavana VH, Peng X, Ge Z, Chen H, Li J, Chen Z, Zhang H, Han L, Du D,

- Creighton CJ, Mills GB; **Cancer Genome Atlas Research Network**, Camargo F, Liang H. Comprehensive Molecular Characterization of the Hippo Signaling Pathway in Cancer. *Cell Rep*. 2018 Oct 30;25(5):1304-1317.e5. doi: 10.1016/j.celrep.2018.10.001. PMID:30380420
138. Slansky JE, **Spellman PT**. Alternative Splicing in Tumors – A Path to Immunogenicity? *N Engl J Med*. 2019 Feb 28;380(9):877-880. doi: 10.1056/NEJMcibr1814237. No abstract available. PMID:30811916
 139. Butler TM, Boniface CT, Johnson-Camacho K, Tabatabaei S, Melendez D, Kelley T, Gray J, Corless CL, **Spellman PT**. Circulating tumor DNA dynamics using patient-customized assays are associated with outcome in neoadjuvantly treated breast cancer. *Cold Spring Harb Mol Case Stud*. 2019 Apr 1;5(2). pii: a003772. doi: 10.1101/mcs.a003772. Print 2019 Apr. PMID:30833418
 140. Rozanov D, Cheltsov A, Nilsen A, Boniface C, Forquer I, Korkola J, Gray J, Tyner J, Tognon CE, Mills GB, **Spellman P**. Targeting mitochondria in cancer therapy could provide a basis for the selective anti-cancer activity. *PLoS One*. 2019 Mar 25;14(3):e0205623. doi: 10.1371/journal.pone.0205623. eCollection 2019. PMID: 30908483
 141. Cassetta L, Fragkogianni S, Sims AH, Swierczak A, Forrester LM, Zhang H, Soong DYH, Cotechini T, Anur P, Lin EY, Fidanza A, Lopez-Yrigoyen M, Millar MR, Urman A, Ai Z, **Spellman PT**, Hwang ES, Dixon JM, Wiechmann L, Coussens LM, Smith HO, Pollard JW. Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets. *Cancer Cell*. 2019 Apr 15;35(4):588-602.e10. doi: 10.1016/j.ccell.2019.02.009. Epub 2019 Mar 28. PMID: 30930117
 142. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CS; **Cancer Genome Atlas Research Network**, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG, Shmulevich I. The Immune Landscape of Cancer. *Immunity*. 2019 Aug 20;51(2):411-412. doi: 10.1016/j.immuni.2019.08.004. No abstract available. PMID: 31433971.
 143. Heskett MB, Sanborn JZ, Boniface C, Goode B, Chapman J, Garg K, Rabban JT, Zaloudek C, Benz SC, **Spellman PT**, Solomon DA, Cho RJ. Multiregion exome sequencing of ovarian immature teratomas reveals 2N near-diploid genomes, paucity of somatic mutations, and extensive allelic imbalances shared across mature, immature, and disseminated components. *Mod Pathol*. 2020 Jun; 33(6): 1193-1206. doi:10.1038/s41379-019-0446-y. Epub 2020 Jan 7. PMID 31911616.
 144. Rubanova Y, Shi R, Harrigan CF, Li R, Wintersinger J, Sahin N, Deshwar A; **PCAWG Evolution and Heterogeneity Working Group**; Morris Q; PCAWG Consortium. Reconstructing evolutionary trajectories of mutation signature activities in cancer using TrackSig. *Nat commun*. 2020 Feb 5;11(1):731. doi: 10.1038/s41467-010-14352-7. PMID: 32024834.
 145. Cmero M, Yuan K, Ong CS, Schröder J; **PCAWG Evolution and Heterogeneity**

- Working Group**, Corcoran NM, Papenfuss T, Hovens CM, Markowitz F, Macintyre G; PCAWG Consortium. Inferring structural variant cancer cell fraction. *Nat Commun*. 2020 Feb 5;11(1):730. doi: 10.1038/s41467-020-14351-8. PMID: 32024845.
146. **ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium**. Pan-Cancer analysis of whole genomes. Version 2. *Nature*. 2020 Feb;578(7793):82-93. doi:10.1038/s41586-020-1969-6. Epub 2020 Feb 5. PMID: 32025007.
 147. Gerstung M, Jolly C, Leshchiner I, D'Entropio SC, Gonzalez S, Rosebrock D, Mitchell TJ, Rubanova Y, Anur P, Yu K, Tarabichi M, Deshwar A, Wintersinger J, Kleinheinz K, Vázquez-García I, Haase K, Jerman L, Sengupta S, Macintyre G, Malikic S, Donmez N, Livitz DG, Cmero M, Demeulemeester J, Schumacher S, Fan Y, Yao X, Lee J, Schlesner M, Boutros PC, Bowtell DD, Zhu H, Getz G, Imielinski M, Beroukhim R, Sahinalp SC, Ji Y, Peifer M, Markowitz F, Mustonen V, Yuan K, Wang W, Morris QD; PCAWG Evolution & Heterogeneity Working Group, **Spellman PT**, Wedge DC, Van Loo P; PCAWG Consortium. The evolutionary history of 2,658 cancers. *Nature*. 2020 Feb;578(7793):122-128. doi:10.1038/s41586-019-1907-7. Epub 2020 Feb 6. PMID: 32025013.
 148. Heskett, MB, Smith LG, **Spellman P**, Thayer MJ. Reciprocal monoallelic expression of ASAR lncRNA genes controls replication timing of human chromosome 6. *RNA*. 2020 Jun;26(6):724-738. doi:10.1261/rna.073114.119. Epub 2020 Mar 6. PMID: 32144193.
 149. Berlow NE, Grasso CS, Quist MJ, Cheng M, Gandour-Edwards R, Hernandez BS, Michalek JE, Ryan C, **Spellman P**, Pal R, Million LS, Renneker M, Keller C. Deep Functional and Molecular Characterization of a High-Risk Undifferentiated Pleomorphic Sarcoma. *Sarcoma*. 2020 May 4;2020:6312480. doi: 10.1155/2020/6312480. PMID: 32565715; PMCID: PMC7285280.
 150. Li CH, Prokopec SD, Sun RX, Yousif F, Schmitz N; PCAWG Tumour Subtypes and Clinical Translation, Boutros PC; **PCAWG Consortium**. Sex differences in oncogenic mutational processes. *Nat Commun*. 2020 Aug 28;11(1):4330. doi: 10.1038/s41467-020-17359-2. PMID: 32859912; PMCID: PMC7455744.
 151. Bailey MH, Meyerson WU, Dursi LJ, Wang LB, Dong G, Liang WW, Weerasinghe A, Li S, Li Y, Kelso S; MC3 Working Group; PCAWG novel somatic mutation calling methods working group, Saksena G, Ellrott K, Wendl MC, Wheeler DA, Getz G, Simpson JT, Gerstein MB, Ding L; **PCAWG Consortium**. Retrospective evaluation of whole exome and genome mutation calls in 746 cancer samples. *Nat Commun*. 2020 Sep 21;11(1):4748. doi: 10.1038/s41467-020-18151-y. Erratum in: *Nat Commun*. 2020 Nov 30;11(1):6232. PMID: 32958763; PMCID: PMC7505971.
 152. Wagner JT, Kim HJ, Johnson-Camacho KC, Kelley T, Newell LF, **Spellman PT**, Ngo TTM. Diurnal stability of cell-free DNA and cell-free RNA in human plasma samples. *Sci Rep*. 2020 Oct 5;10(1):16456. doi: 10.1038/s41598-020-73350-3. PMID: 33020547; PMCID: PMC7536441.
 153. Zhao W, Li J, Chen MM, Luo Y, Ju Z, Nesser NK, Johnson-Camacho K, Boniface CT, Lawrence Y, Pande NT, Davies MA, Herlyn M, Muranen T, Zervantonakis IK, von Euw E, Schultz A, Kumar SV, Korkut A, **Spellman PT**, Akbani R, Slamon DJ, Gray JW, Brugge JS, Lu Y, Mills GB, Liang H. Large-Scale Characterization of Drug Responses of Clinically Relevant Proteins in Cancer Cell Lines. *Cancer Cell*. 2020 Oct 26;S1535-6108(20)30539-0. doi: 10.1016/j.ccell.2020.10.008. Epub ahead of print. PMID: 33157050.

154. Naffar-Abu Amara S, Kuiken HJ, Selfors LM, Butler T, Leung ML, Leung CT, Kuhn EP, Kolarova T, Hage C, Ganesh K, Panayiotou R, Foster R, Rueda BR, Aktipis A, **Spellman P**, Ince TA, Xiu J, Oberley M, Gatalica Z, Navin N, Mills GB, Bronson RT, Brugge JS. Transient commensal clonal interactions can drive tumor metastasis. *Nat Commun*. 2020 Nov 16;11(1):5799. doi: 10.1038/s41467-020-19584-1. PMID: 33199705; PMCID: PMC7669858.
155. Boniface C, Deig C, Halsey C, Kelley T, Heskett MB, Thomas CR Jr, **Spellman PT**, Nabavizadeh N. The Feasibility of Patient-Specific Circulating Tumor DNA Monitoring throughout Multi-Modality Therapy for Locally Advanced Esophageal and Rectal Cancer: A Potential Biomarker for Early Detection of Subclinical Disease. *Diagnostics (Basel)*. 2021 Jan 5;11(1):73. doi: 10.3390/diagnostics11010073. PMID: 33466369
156. Matsumoto T, Wakefield L, Peters A, Peto M, **Spellman P**, Grompe M. Proliferative polyploid cells give rise to tumors via ploidy reduction. *Nat Commun*. 2021 Jan 28;12(1):646. doi: 10.1038/s41467-021-20916-y. PMID: 33510149
157. Carrot-Zhang J, Yao X, Devarakonda S, Deshpande A, Damrauer JS, Silva TC, Wong CK, Choi HY, Felau I, Robertson AG, Castro MAA, Bao L, Rheinbay E, Liu EM, Trieu T, Haan D, Yau C, Hinoue T, Liu Y, Shapira O, Kumar K, Mungall KL, Zhang H, Lee JJ, Berger A, Gao GF, Zhitomirsky B, Liang WW, Zhou M, Moorthi S, Berger AH, Collisson EA, Zody MC, Ding L, Cherniack AD, Getz G, Elemento O, Benz CC, Stuart J, Zenklusen JC, Beroukhi R, Chang JC, Campbell JD, Hayes DN, Yang L, Laird PW, Weinstein JN, Kwiatkowski DJ, Tsao MS, Travis WD, Khurana E, Berman BP, Hoadley KA, Robine N; **TCGA Research Network**, Meyerson M, Govindan R, Imielinski M. Whole-genome characterization of lung adenocarcinomas lacking the RTK/RAS/RAF pathway. *Cell Rep*. 2021 Feb 2;34(5):108707. doi: 10.1016/j.celrep.2021.108707. PMID: 33535033
158. D'Entrop SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, Yu K, Rubanova Y, Macintyre G, Demeulemeester J, Vázquez-García I, Kleinheinz K, Livitz DG, Malikic S, Donmez N, Sengupta S, Anur P, Jolly C, Cmero M, Rosebrock D, Schumacher SE, Fan Y, Fittall M, Drews RM, Yao X, Watkins TBK, Lee J, Schlesner M, Zhu H, Adams DJ, McGranahan N, Swanton C, Getz G, Boutros PC, Imielinski M, Beroukhi R, Sahinalp SC, Ji Y, Peifer M, Martincorena I, Markowitz F, Mustonen V, Yuan K, Gerstung M, **Spellman PT**, Wang W, Morris QD, Wedge DC, Van Loo P; PCAWG Evolution and Heterogeneity Working Group and the PCAWG Consortium. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell*. 2021 Apr 15;184(8):2239-2254.e39. doi: 10.1016/j.cell.2021.03.009. Epub 2021 Apr 7. PMID: 33831375
159. Janiszewska M, Stein S, Metzger Filho O, Eng J, Kingston NL, Harper NW, Rye IH, Alečković M, Trinh A, Murphy KC, Marangoni E, Cristea S, Oakes B, Winer EP, Krop IE, Russnes HG, **Spellman PT**, Bucher E, Hu Z, Chin K, Gray JW, Michor F, Polyak K. The impact of tumor epithelial and microenvironmental heterogeneity on treatment responses in HER2+ breast cancer. *JCI Insight*. 2021 Jun 8;6(11):e147617. doi: 10.1172/jci.insight.147617. PMID: 33886505
160. Babur Ö, Luna A, Korkut A, Durupinar F, Siper MC, Dogrusoz U, Vaca Jacome AS, Peckner R, Christianson KE, Jaffe JD, **Spellman PT**, Aslan JE, Sander C, Demir E. Causal interactions from proteomic profiles: Molecular data meet pathway knowledge. *Patterns (N Y)*. 2021 May 12;2(6):100257. doi: 10.1016/j.patter.2021.100257.

- eCollection 2021 Jun 11. PMID: 34179843
161. Minnier J, Rajeevan N, Gao L, Park B, Pyarajan S, **Spellman P**, Haskell SG, Brandt CA, Luoh SW. Polygenic Breast Cancer Risk for Women Veterans in the Million Veteran Program. *JCO Precis Oncol*. 2021 Jul 21;5:PO.20.00541. doi: 10.1200/PO.20.00541. eCollection 2021 Jul. PMID: 34381935
 162. Creason A, Haan D, Dang K, Chiotti KE, Inkman M, Lamb A, Yu T, Hu Y, Norman TC, Buchanan A, van Baren MJ, Spangler R, Rollins MR, **Spellman PT**, Rozanov D, Zhang J, Maher CA, Caloian C, Watson JD, Uhrig S, Haas BJ, Jain M, Akeson M, Ahsen ME; SMC-RNA Challenge Participants, Stolovitzky G, Guinney J, Boutros PC, Stuart JM, Ellrott K. A community challenge to evaluate RNA-seq, fusion detection, and isoform quantification methods for cancer discovery. *Cell Syst*. 2021 Aug 18;12(8):827-838.e5. doi: 10.1016/j.cels.2021.05.021. Epub 2021 Jun 18. PMID: 34146471
 163. Sun D, Guan X, Moran AE, Wu LY, Qian DZ, Schedin P, Dai MS, Danilov AV, Alumkal JJ, Adey AC, **Spellman PT**, Xia Z. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat Biotechnol*. 2021 Nov 11. doi: 10.1038/s41587-021-01091-3. PMID: 34764492
 164. Eksi SE, Chitsazan A, Sayar Z, Thomas GV, Fields AJ, Kopp RP, **Spellman PT**, Adey AC. Epigenetic loss of heterogeneity from low to high grade localized prostate tumours. *Nat Commun*. 2021 Dec 15;12(1):7292. doi: 10.1038/s41467-021-27615-8.
 165. Lindner JR, Davidson BP, Song Z, Maier CS, Minnier J, Stevens JF, Ferencik M, Taqui S, Belcik JT, Moccetti F, Layoun M, **Spellman P**, Turker MS, Tavori H, Fazio S, Raber J, Bobe G. Plasma Lipidomic Patterns in Patients with Symptomatic Coronary Microvascular Dysfunction. *Metabolites*. 2021 Sep 22;11(10):648. doi: 10.3390/metabo11100648.
 166. Labrie M, Li A, Creason A, Betts C, Keck J, Johnson B, Sivagnanam S, Boniface C, Ma H, Blucher A, Chang YH, Chin K, Vuky J, Guimaraes AR, Downey M, Lim JY, Gao L, Siex K, Parmar S, Kolodzie A, **Spellman PT**, Goecks J, Coussens LM, Corless CL, Bergan R, Gray JW, Mills GB, Mitri ZI. Multiomics analysis of serial PARP inhibitor treated metastatic TNBC inform on rational combination therapies. *NPJ Precis Oncol*. 2021 Oct 19;5(1):92. doi: 10.1038/s41698-021-00232-w. PMID: 34667258
 167. Johnson BE, Creason AL, Stommel JM, Keck JM, Parmar S, Betts CB, Blucher A, Boniface C, Bucher E, Burlingame E, Camp T, Chin K, Eng J, Estabrook J, Feiler HS, Heskett MB, Hu Z, Kolodzie A, Kong BL, Labrie M, Lee J, Leyshock P, Mitri S, Patterson J, Riesterer JL, Sivagnanam S, Somers J, Sudar D, Thibault G, Weeder BR, Zheng C, Nan X, Thompson RF, Heiser LM, **Spellman PT**, Thomas G, Demir E, Chang YH, Coussens LM, Guimaraes AR, Corless C, Goecks J, Bergan R, Mitri Z, Mills GB, Gray JW. An omic and multidimensional spatial atlas from serial biopsies of an evolving metastatic breast cancer. *Cell Rep Med*. 2022 Feb 15;3(2):100525. doi:10.1016/j.xcrm.2022.100525. eCollection 2022 Feb 15. PMID: 35243422
 168. Roskams-Hieter B, Kim HJ, Anur P, Wagner JT, Callahan R, Spiliotopoulos E, Kirschbaum CW, Civitci F, **Spellman PT**, Thompson RF, Farsad K, Naugler WE, Ngo TTM. Plasma cell-free RNA profiling distinguishes cancers from pre-malignant conditions in solid and hematologic malignancies. *NPJ Precis Oncol*. 2022 Apr 25;6(1):28. doi: 10.1038/s41698-022-00270-y.
 169. Heskett MB, Vouzas AE, Smith LG, Yates PA, Boniface C, Bouhassira EE, **Spellman PT**, Gilbert DM, Thayer MJ. Epigenetic control of chromosome-associated lncRNA

- genes essential for replication and stability. *Nat Commun.* 2022 Oct 22;13(1):6301. doi: 10.1038/s41467-022-34099-7.
170. Gurun B, Horton W, Murugan D, Zhu B, Leyshock P, Kumar S, Byrne KT, Vonderheide RH, Margolin AA, Mori M, Spellman PT, Coussens LM, Speed TP. An open protocol for modeling T Cell Clonotype repertoires using TCR β CDR3 sequences. *BMC Genomics.* 2023 Jun 26;24(1):349. doi: 10.1186/s12864-023-09424-z.
 171. O'Brien TD, Potter AB, Driscoll CC, Goh G, Letaw JH, McCabe S, Thanner J, Kulkarni A, Wong R, Medica S, Week T, Buitrago J, Larson A, Camacho KJ, Brown K, Crist R, Conrad C, Evans-Dutson S, Lutz R, Mitchell A, Anur P, Serrato V, Shafer A, Marriott LK, Hamman KJ, Mulford A, Wiszniewski W, Sampson JE, Adey A, O'Roak BJ, Harrington CA, Shannon J, Spellman PT, Richards CS. Population screening shows risk of inherited cancer and familial hypercholesterolemia in Oregon. *Am J Hum Genet.* 2023 Aug 3;110(8):1249-1265. doi: 10.1016/j.ajhg.2023.06.014. Epub 2023 Jul 27.
 172. Li Z, Metzger Filho O, Viale G, dell'Orto P, Russo L, Goyette MA, Kamat A, Yardley DA, Gupta Abramson V, Arteaga CL, Spring LM, Chiotti K, Halsey C, Waks AG, King TA, Lester SC, Bellon JR, Winer EP, Spellman PT, Krop IE, Polyak K. HER2 heterogeneity and treatment response-associated profiles in HER2-positive breast cancer in the NCT02326974 clinical trial. *J Clin Invest.* 2024 Feb 1;134(7):e176454. doi: 10.1172/JCI176454.
 173. Zhang Z, Shafer A, Johnson-Camacho K, Adey A, Anur P, Brown KA, Conrad C, Crist R, Farris PE, Harrington CA, Marriott LK, Mitchell A, O'Roak B, Serrato V, Richards CS, Spellman PT, Shannon J. Novel recruitment approaches and operational results for a statewide population Cohort for cancer research: The Healthy Oregon Project. *J Clin Transl Sci.* 2024 Jan 19;8(1):e32. doi: 10.1017/cts.2024.9. eCollection 2024.
 174. Marín-Aguilera M, Jares P, Sanfeliu E, Villacampa G, Hernández-Lllán E, Martínez-Puchol AI, Shankar S, González-Farré B, Waks AG, Brasó-Maristany F, Pardo F, Manning DK, Abery JA, Curaba J, Moon L, Gordon O, Galván P, Wachirakantapong P, Castillo O, Nee CM, Blasco P, Senevirathne TH, Sirenko V, Martínez-Sáez O, Aguirre A, Krop IE, Li Z, Spellman P, Metzger Filho O, Polyak K, Michaels P, Puig-Butillé JA, Vivancos A, Matito J, Buckingham W, Perou CM, Villagrasa-González P, Prat A, Parker JS, Paré L. Analytical validation of HER2DX genomic test for early-stage HER2-positive breast cancer. *ESMO Open.* 2024 Mar;9(3):102903. doi: 10.1016/j.esmoop.2024.102903. Epub 2024 Mar 6.
 175. Vermehren-Schmaedick A, Peto M, Wagoner W, Chiotti KE, Ramsey E, Wang X, Rakshe S, Minnier J, Sears R, Spellman P, Luoh SW. Mutation of SIVA, a candidate metastasis gene identified from clonally related bilateral breast cancers, promotes breast cancer cell spread in vitro and in vivo. *PLoS One.* 2024 May 9;19(5):e0302856. doi: 10.1371/journal.pone.0302856. eCollection 2024.

B. RESEARCH PAPERS – PEER REVIEWED (IN PRESS)

NA

C. RESEARCH PAPERS – PEER REVIEWED (SUBMITTED)

CHAPTERS

1. **P. T. Spellman.** Cluster Analysis and Display. DNA Microarrays, A Molecular Cloning Manual, Bowtell D. and Sambrook J., Ed., *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor. 569-581, 2002.
2. **P.T. Spellman**, L.M. Heiser, J.W. Gray. Chapter 27 Breast Cancer Genomics. Section VI Pathology and Biological Markers of Invasive Breast Cancer. In: Diseases of the Breast, 4th Edition (Harris, Lippman, Morrow & Osborne, eds.). pp. 363-373. 2009
3. **P.T. Spellman**, J.F. Costello, J.W. Gray. Molecular Basis of Cancer, Mendelsohn J., Ed., *Elsevier*. 2009

LETTERS TO THE EDITOR

1. C. A. Ball, G. Sherlock, H. Parkinson, P. Rocca-Sera, C. Brooksbank, H. C. Causton, D. Cavalieri, T. Gaasterland, P. Hingamp, F. Holstege, M. Ringwald, **P. Spellman**, C. J. Stoeckert, Jr., J. E. Stewart, R. Taylor, A. Brazma and J. Quackenbush. Standards for microarray data. *Science*. 298: 539, 2002
2. C. A. Ball, G. Sherlock, H. Parkinson, P. Rocca-Sera, C. Brooksbank, H. C. Causton, D. Cavalieri, T. Gaasterland, P. Hingamp, F. Holstege, M. Ringwald, **P. Spellman**, C. J. Stoeckert, Jr., J. E. Stewart, R. Taylor, A. Brazma and J. Quackenbush. The underlying principles of scientific publication. *Bioinformatics*. 18: 1409, 2002
3. C. A. Ball, A. Brazma, H. Causton, S. Chervitz, R. Edgar, P. Hingamp, J.C. Matese, H. Parkinson, J. Quackenbush, M. Ringwald, S. Sansone, G. Sherlock, **P. Spellman**, C. Stoeckert, Y. Tateno, R. Taylor, J. White, N. Winegarden. Submission of Microarray Data to Public Repositories. *PLOS Biology*. 2: e317, 2004
4. Ball C, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, Matese JC, Parkinson H, Quackenbush J, Ringwald M, Sansone SA, Sherlock G, **Spellman P**, Stoeckert C, Tateno Y, Taylor R, White J, Winegarden N; MGED Society. Standards for microarray data: an open letter. *Environ Health Perspect*. 112(12):A666-7, Aug 2004
5. Quackenbush J, Stoeckert C, Ball C, Brazma A, Gentleman R, Huber W, Irizarry R, Salit M, Sherlock G, **Spellman P**, Winegarden N. Top-down standards will not serve systems biology. *Nature*. 2; 440(7080):24, Mar 2006

REVIEWS

1. **P. T. Spellman.** The future of publishing microarray data. *Briefings in Bioinformatics*. 2: 316-8, 2001
2. **P. T. Spellman.** Cluster Analysis and Display. DNA Microarrays, A Molecular Cloning Manual, Bowtell D. and Sambrook J., Ed., *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor. 569-581, 2002
3. **P. T. Spellman** and G. Sherlock. Final words: cell age and cell cycle are unlinked. *Trends Biotechnology*. 6: 270-3, 2004
4. **P. T. Spellman** and G. Sherlock. Reply: whole-culture synchronization – effective tools for cell cycle studies. *Trends Biotechnology*. 6: 277-8, 2004
5. **P. Spellman.** A status report on MAGE. *Bioinformatics*. 21:3459-60, 2005
6. **P. Spellman** and J. Gray. A new treasure in the breast cancer gene hunt. *Nat. Med*. 2011
7. C.L. Corless and **P.T. Spellman.** Tackling formalin-fixed, paraffin-embedded tumor tissue with next-generation sequencing. *Cancer Discovery*. 2012
8. **P.T. Spellman** and J.W. Gray. Detecting cancer by monitoring circulating tumor DNA.

Nat Med. 2014

9. Butler TM, **Spellman PT**, Gray J. Circulating-tumor DNA as an early detection and diagnostic tool. *Curr Opin Genet Dev*. 2017 Feb;42:14-21. doi: 10.1016/j.gde.2016.12.003. Epub 2017 Jan 23. Review.
10. Heskett MB, **Spellman PT**, Thayer MJ. Differential Allelic Expression among Long Non-Coding RNAs. *Noncoding RNA*. 2021 Oct 22;7(4):66. doi: 10.3390/ncrna7040066.
11. Boniface CT, **Spellman PT**, Blood, Toil, and Taxoteres: Biological Determinates of Treatment-Induce ctDNA Dynamics for Interpreting Tumor Response. *Pathol Oncol Res*. 2022 May 19;28:1610103. doi: 10.3389/pore.2022.1610103.
12. Goldman E, **Spellman PT**, Agarwal A. Defining clonal hematopoiesis of indeterminate potential: evolutionary dynamics and detection under aging and inflammation. *Goldman EA, Spellman PT, Agarwal A. Cold Spring Harb Mol Case Stud*. 2023 May 9;9(2):a006251. doi: 10.1101/mcs.a006251. Print 2023 Apr.

Abstracts

1. Systems Biology: Genomic Approaches to Transcriptional Regulation, March 6-9, 2003, "Characterization of transcription neighborhoods in *Drosophila*"
2. MGED7, September 2004, Toronto, Canada. "Molecular Inversion Probes for Copy Number Analysis"
3. Conference speaker, ICBP Mouse Models, Dec 1-3, 2006, Bethesda, MD. "Signaling Networks and Therapeutic Agent Response in Breast Cancer Cell Lines"
4. MGED13, September 2009, Phoenix, AZ. "Omic approaches to predictive markers"

Other

Invited Lectures, Conference Presentations or Professorships:

Appendix B

LIST OF MATERIALS CONSIDERED

Exhibit	Description
1001	U.S. Patent No. 11,768,200
1004	P. Parameswaran, et al., <i>A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing</i> , 35 NUCLEIC ACIDS RSCH. 19(e130) (2007).
1005	G.B. Gloor, et al., <i>Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products</i> , Cornell University Library arXiv:1007.5075v1 (2010).
1006	D.R. Bentley, et al., <i>Accurate whole human genome sequencing using reversible terminator chemistry</i> , 456 NATURE 6, 53-59 (2008).
1007	File History of U.S. Patent No. 11,768,200
1009	F. Sanger, et al., <i>DNA sequencing with chain-terminating inhibitors</i> , 74 PROC. NAT'L. ACAD. SCI. USA 12, 5463-5467 (1977).
1010	M. Kircher & J. Kelso, <i>High-throughput DNA sequencing – concepts and limitations</i> , 32 BIOESSAYS 6, 524-536 (2010).
1011	M.L. Metzker, <i>Sequencing technologies—the next generation</i> , 11 NATURE REVIEWS GENETICS 31-46 (2010).
1012	J. Shendure & H. Ji, <i>Next-generation DNA sequencing</i> , 26 NATURE BIOTECH. 10, 1135-1145 (2008).
1013	G. Wiley, et al., <i>Methods for generating shotgun and mixed shotgun/paired-end libraries for the 454 DNA sequencer</i> , 61 CURRENT PROTOCOLS IN HUMAN GENETICS 1, 18.1.1-18.1.21 (2009).
1014	M. Meyer, et al., <i>Parallel tagged sequencing on the 454 platform</i> , 3 NATURE PROTOCOLS 2, 267-278 (2008).
1015	M. Meyer & M. Kircher, <i>Illumina sequencing library preparation for highly multiplexed target capture and sequencing</i> , 2010, COLD SPRING HARBOR PROTOCOL 6 (2010).
1016	U.S. Application Publication No. 2012/0028814
1017	S.J.G. Lahr & L.A. Katz, <i>Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies</i>

	<i>using a new-generation high-fidelity DNA polymerase</i> , 47 BIOTECHNIQUES 4, 857-866 (2009).
1018	A. Meyerhans, et al., <i>DNA recombination during PCR</i> , 18 NUCLEIC ACIDS RSCH. 7, 1687-1691 (1990).
1019	S. Paabo, et al., <i>DNA damage promotes jumping between templates during enzymatic amplification</i> , 265 J. OF BIOL. CHEM. 8, 4718-4721 (1990).
1020	S.J. Odelberg, et al., <i>Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I</i> , 23 NUCLEIC ACIDS RSCH. 11, 2049-2057 (1995).
1021	D.W. Craig, et al., <i>Identification of genetic variants using bar-coded multiplexed sequencing</i> , 5 NATURE METHODS 10, 887-893 (2008).
1022	R. Li, et al., <i>SNP detection for massively parallel whole-genome resequencing</i> , 19 GENOME RSCH. 1124-1132 (2009).
1023	M. Choi, et al., <i>Genetic diagnosis by whole exome capture and massively parallel DNA sequencing</i> , 106 PNAS 45, 19096-19101 (2009).
1024	G.B. Gloor, et al., <i>Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products</i> , 5 PLOS ONE 10(e15406) (2010).
1025	<i>DNA Sequencing With Solexa® Technology</i> , Technology Spotlight, Illumina® Systems & Software, Pub. No. 770-2007-002 (2007).
1026	File History of U.S. Patent No. 9,163,281
1027	Illumina Press Release, <i>Illumina Announces New TruSeq™ SBS Sequencing Kit for the Genome Analyzer</i> , ILLUMINA (July 26, 2010), accessed at https://emea.illumina.com/cmopany/news-center/press-releases/2010/1451638.html on June 1, 2024.
1028	<i>Genome Sequencer 20 System First to the Finish</i> , Roche Diagnostics®, Pub. No. 04866452990 (2006).
1029	J. Binladen, et al., <i>The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing</i> , 2 PLOS ONE 2:e197 (2007).
1030	U.S. Patent No. 9,163,281
1031	Bruce Alberts, et al., <i>Molecular Biology of the Cell</i> 4, GARLAND SCIENCE (2002).

1032	<i>Polymerase Chain Reaction (PCR)</i> , NATIONAL HUMAN GENOME RESEARCH INSTITUTE (updated June 26, 2024), https://www.genome.gov/genetics-glossary/Polymerase-Chain-Reaction .
1033	M. Dai, et al., <i>NGSQC: cross-platform quality analysis pipeline for deep sequencing data</i> , 11 BMC GENOMICS 11(Suppl 4):S7 (2010).
1034	S. Creer, et al., <i>Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises</i> , 19 MOLECULAR ECOLOGY (Suppl. 1) 4-20 (2010).
1035	The Cancer Genome Atlas Research Network, <i>Comprehensive genomic characterization defines human glioblastoma genes and core pathways</i> , 455 NATURE 23 (2008).
1038	P. Maughan, et al., <i>Single-Nucleotide Polymorphism Genotyping in Mapping Populations via Genomic Reduction and Next-Generation Sequencing: Proof of Concept</i> , 3 PLANT GENOME (3) 166-178 (2010).