

Case 2024-00964
U.S. Patent No. 11,041,852

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ILLUMINA INC.,
Petitioner

v.

MOLECULAR LOOP BIOSCIENCES, INC.,
Patent Owner

Case IPR2024-00964

U.S. Patent No. 11,041,852
TITLE: Methods for Maintaining the Integrity and Identification of a Nucleic Acid
Template in a Multiplex Sequencing Reaction
Issue Date: June 22, 2021

DECLARATION OF VIKTOR ADALSTEINSSON, PH.D.

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	4
II. BASES FOR OPINIONS	4
III. MATERIALS REVIEWED	5
IV. EDUCATION AND EXPERIENCE.....	5
V. LEGAL STANDARDS	8
A. Claim Construction.....	9
B. Obviousness.....	9
VI. LEVEL OF ORDINARY SKILL IN THE ART.....	12
VII. CLAIM CONSTRUCTION	15
VIII. TECHNICAL BACKGROUND	21
IX. PROSECUTION HISTORY OF THE '852 PATENT.....	32
X. ILLUMINA'S ASSERTED PRIOR ART.....	33
A. Parameswaran.....	33
B. Gloor.....	45
C. Bentley.....	53
XI. GROUND I: THE CHALLENGED CLAIMS ARE NOT OBVIOUS OVER PARAMESWARAN AND GLOOR.....	58
A. A POSA Would Not Be Motivated To Combine Parameswaran and Gloor.....	58
1. Dr. Spellman Does Not Consider The Full Context Of The References And The State Of The Art.....	58

2.	Dr. Spellman’s Reasons To Combine Parameswaran and Gloor Fail To Show A Motivation To Combine Unique Dual Indexes With Illumina Sequencers.	64
i.	Dr. Spellman Does Not Show That Efficiency And Cost Considerations Would Have Motivated The Proposed Combination.	64
ii.	Dr. Spellman Does Not Show That A Desire To “Reduce Sequencing Error” Would Have Motivated The Proposed Combination.	67
iii.	Dr. Spellman Does Not Show That Parameswaran’s Speculation Regarding Other Sequencers Would Have Motivated The Proposed Combination.....	90
B.	A POSA Would Not Have Reasonably Expected Success In Combining Parameswaran With Gloor	91
C.	Dr. Spellman Fails To Show That The Combination Discloses Or Renders Obvious “Discarding Sequence Reads Obtained In Said Sequencing Step That Contain Said Combination Of Identifier Sequences That Are Different Than Any Of Said Distinct Pairs, Thereby To Reduce Cross-Over Error.”	96
XII.	GROUND 2: THE CHALLENGED CLAIMS ARE NOT OBVIOUS OVER PARAMESWARAN AND BENTLEY	104
A.	A POSA Would Not Be Motivated To Combine Parameswaran And Bentley.....	104
B.	A POSA Would Not Have Reasonably Expected Success In Combining Parameswaran and Bentley.	105
C.	Dr. Spellman Fails To Show That The Combination Discloses Or Renders Obvious “Discarding Sequence Reads Obtained In Said Sequencing Step That Contain Said Combination Of Identifier Sequences That Are Different Than Any Of Said Distinct Pairs, Thereby To Reduce Cross-Over Error.”	106
XIII.	APPENDIX A: THE CHALLENGED CLAIMS OF THE ’852 PATENT	108

I. INTRODUCTION

1. I, Viktor Adalsteinsson, Ph.D., a resident of Acton, Massachusetts, over 18 years of age, hereby declare as follows:

- I have personal knowledge of all of the matters about which I testify in this declaration.
- I have been retained on behalf of the Patent Owner Molecular Loop to provide my technical opinions and testimony about U.S. Patent No. 11,041,852 (the “’852 Patent”) and its Claims 1-8 (the “Challenged Claims”). The full text of the Challenged Claims appears in Appendix A to this declaration.
- I am being compensated for my work in this proceeding and receiving reimbursement for expenses incurred in the course of my work. My compensation is not contingent in any way on either the opinions I have reached or the outcome of this case.

II. BASES FOR OPINIONS

2. I have reviewed and considered the documents and other materials listed below in Section III in light of my specialized knowledge provided by my education, training, research, and experience, as summarized in Section IV and described in detail in my CV attached to this declaration. My analysis of those materials, combined with the

specialized knowledge that I have obtained over the course of my education and career, form the bases for my opinions in this declaration.

III. MATERIALS REVIEWED

3. In forming the opinions set forth in this declaration, I have relied upon:
 - my experience and knowledge in the field of biochemistry, molecular biology, genetics, advanced genomic technologies, cancer diagnostics, and liquid biopsy;
 - U.S. Patent Nos. 11,041,852 and 11,768,200;
 - the prosecution history of U.S. Patent No. 11,041,852;
 - the prosecution history of U.S. Patent No. 11,768,200;
 - Illumina's Petition for *inter partes* Review and the materials cited therein;
 - the declaration of Dr. Paul Spellman (EX1003) and the materials cited therein, as well as the transcript of the April 9, 2025 deposition testimony of Dr. Spellman (EX2002); and
 - all other materials cited herein.

IV. EDUCATION AND EXPERIENCE

4. I am currently the Director of the Gerstner Center for Cancer Diagnostics at the Broad Institute of MIT and Harvard. I also run my own laboratory at the Broad Institute focused on developing novel

genomic technologies to improve cancer diagnosis and monitoring and to inform more precise cancer therapy.

5. I earned a Bachelor of Science in Chemical Engineering in 2010 from the Schreyer Honors College at Pennsylvania State University. I then earned my Ph.D. in Chemical Engineering in 2015 from the Massachusetts Institute of Technology, during which I studied the genome sequencing and phenotypic analysis of single cells in cancer. Following my Ph.D., I completed a Postdoctoral research position in Chemical Engineering at Massachusetts Institute of Technology.
6. I have ten years of experience as a principal investigator overseeing assay development and translational research studies, in collaboration with investigators at Harvard Medical School. Since joining the Broad Institute, I established and led the Blood Biopsy Team in 2015. The Blood Biopsy Team is a multi-institutional effort spanning Broad, MIT, Harvard, and the Harvard-affiliated hospitals, to enable routine profiling of cancer genomes from blood samples. In 2019, I also helped establish and lead the Gerstner Center for Cancer Diagnostics. The Gerstner Center is an innovation hub for new advances in cancer diagnostics at the Broad Institute.

7. I have 15 years of experience researching Cancer Genomics and Cancer Diagnostics. I have developed and applied nucleic acid sequencing and analysis methods to over 15,000 liquid biopsy samples, including an ultrasensitive test for minimal residual disease. My research involved developing some of the first methods to isolate and perform whole-exome and whole-genome sequencing of single circulating tumor cells. I also developed workflows for comprehensive sequencing of circulating tumor DNA, which have been used in large-scale studies to profile cancer drug resistance. In addition, I developed methods for highly accurate nucleic acid sequencing which are being used to study somatic mosaicism in normal tissues, to detect cancer mutations from liquid biopsies, and to identify mutation signatures.
8. I have mentored or co-mentored over 20 emerging scientists who have gone to prestigious graduate or postgraduate programs or started their own labs. I have also served as a grant reviewer for the National Institutes of Health, as a peer reviewer for journals such as Nature Biotechnology, Proceedings of the National Academies of Sciences, Nature Genetics, among others, as a guest lecturer in courses and seminar series, and as a presenter, panelist, or session organizer at

major conferences such as the American Association for Cancer Research.

9. I have published over 45 papers with over 5,000 citations in the fields of cancer genomics, cancer diagnostics, and biotechnology, including in journals such as Science, Journal of Clinical Oncology, Nature Biotechnology, Nature Genetics, Nature Biomedical Engineering, Nature Communications, and Annals of Oncology. My research has been featured by news and media outlets such as WBUR, PBS, and STAT, among others. I have been awarded over \$25M in philanthropic support and federal grants, including from the National Institutes of Health, for my work on nucleic acid sequencing. I was honored by the MIT Technology Review as one of its 35 Innovators Under 35 and by Clinical OMICs as one of its Pioneers Under 40.

V. LEGAL STANDARDS

10. Counsel for Molecular Loop have helped me to understand the applicable legal standards for this case. I have described my understanding of those standards below. I am not an attorney. I do not have any formal training in patent law. My opinions are based on my understanding of the legal principles as I describe them below.

A. Claim Construction

11. I understand that the court decides claim construction issues, which are a matter of law. I understand the claim construction inquiry as follows. I understand the claim language itself to be the first focus, followed by the language in the patent's intrinsic record, including the abstract, specification, figures, and prosecution history. Counsel for Molecular Loop has informed me that communications between the Applicant and the United States Patent and Trademark Office (USPTO) make up the prosecution history. I understand extrinsic evidence, like dictionaries and publications, to be helpful in explaining the meaning of claim terms, but that claim terms should be construed according to the intrinsic evidence.

B. Obviousness

12. I understand that a patent claim will be found obvious and invalid only if the subject matter would have been obvious to a person of ordinary skill in the art ("POSA") at the time of the invention when comparing it to the prior art. I also understand that a claim will be invalid if two or more prior art references disclose in combination every claim limitation such that it would have made the invention obvious to a POSA at the time it was made.

13. I understand that determining whether an invention is obvious requires making decisions on several fact issues, but that the ultimate determination is a question of law. I understand the relevant factual questions to require: (1) determining the content and scope of the prior art, (2) considering as a whole how the claimed invention is different from the prior art, (3) determining the level of ordinary skill that a POSA would have had at the time of the invention, and (4) considering whether the invention contributed to certain objective, contemporaneous responses by others at the time, such as by considering whether the invention had commercial success, fulfilled a long-felt need, led to unexpected results, or was not arrived at by others.
14. I understand that determining whether an invention is obvious requires considering the state of the art at the time of the invention. Thus, I took care to avoid importing any hindsight biases into my obviousness analysis. I also understand that it is improper to use the invention as a model for putting together pieces of the prior art to form the invention. For example, if, at the time of the invention, a POSA knew of the problem and was instructed to make the patented invention, they would almost always be able to make the invention. So, in performing my obviousness analysis, I have considered whether the inventive

contribution included defining the problem in a new way or discovering that a problem existed. I am aware that determining obviousness requires showing, by a preponderance of the evidence, that a POSA would have both recognized the specific problem identified by the inventor and found it obvious to use the inventor's methods as a solution to that problem.

15. I also understand that the party challenging the patent has the burden of proving an invention is obvious. Counsel for Molecular Loop has informed me that this burden requires showing that a POSA would have had a reason to combine the elements in the manner claimed with a reasonable expectation of success. I understand that, although a prior art reference may expressly or inherently disclose a claim limitation, a claim is not proven to be obvious just by showing that each of the elements was independently known in the prior art. Instead, the patent challenger must provide a reason why it would have been obvious to change the prior art to arrive at the invention.
16. I understand that any conclusory statements are insufficient to show obviousness. I have considered whether Petitioner and Dr. Spellman have provided support to explain why they arrive at a conclusion that the invention is obvious.

17. I understand that a POSA would not modify prior art if the potential modification would have made the prior art unfit for the intended purpose. I am also aware that a POSA would have needed a motivation to combine the references. For example, I have considered whether combining several references to reach a conclusion of obviousness would have required substantial reconstruction of the elements and a change in the fundamental principles that the prior art was designed to operate on.
18. I am also aware that, for something to be prior art for obviousness, one must consider whether it is analogous to the claimed invention such that a POSA would have reasonably considered it and applied its teachings when seeking a solution to the problem identified by the inventor as in need of a solution.

VI. LEVEL OF ORDINARY SKILL IN THE ART

19. I understand Petitioner and Dr. Spellman's position is that "[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.” Illumina’s Petition (“Pet.”), 17; EX1003 (Spellman Declaration), ¶¶ 19-20.

20. Counsel for Molecular Loop has informed me that a technical expert need not have acquired the requirements for a POSA before the time of the invention to be qualified to offer a reliable opinion on the perspective of a POSA if they possess the technical knowledge and ability to do so. I believe I am amply qualified to offer opinions on the perspective of a POSA in this case. I received my Ph.D. in Chemical Engineering in 2015. In the work I did in the years leading up to receiving my Ph.D., I completed research in a lab in the Department of Chemical Engineering at the Massachusetts Institute of Technology. During my Ph.D. studies, I developed methods to isolate, perform whole-genome amplification, conduct whole-exome and whole-genome sequencing of single circulating tumor cells, and derive accurate consensus variant calls that reflected matched tumor biopsies. This work was published in Nature Biotechnology in 2014. I also developed methods to calibrate genomic and allelic bias in single-cell sequencing datasets derived from whole-genome amplification methods such as multiple displacement amplification. This work was published in Nature Communications in 2015. For these studies, I used

Illumina sequencing technologies including Illumina kits for sequencing library construction and non-unique dual index barcoding.

I wrote my dissertation on genomic sequencing and phenotypic analysis of single cells in cancer which described my work on both single-cell sequencing and functional profiling of single tumor cells using arrays of sub-nanoliter wells to characterize secreted cytokines and cell phenotypes.

21. After receiving my Ph.D., I completed a postdoctoral research position in chemical engineering in the same MIT laboratory where I performed research for my Ph.D. Since then, I have continued my research in high-throughput sequencing technologies, including through my experience researching cancer genomics and diagnostics to further refine methods for isolation and sequencing of single circulating tumor cells. This research has led me to develop workflows for comprehensive sequencing of circulating tumor DNA and methods for highly accurate nucleic acid sequencing, which are now used to detect minimal residual disease, study somatic mosaicism, detect cancer mutations, and identify mutation signatures. In that work, I have gained an intimate knowledge of the detailed workings of high-throughput next-generation sequencing technologies, both for the technologies available today and

the technologies that were available in 2010 when NGS was first emerging. In total I have acquired over 15 years of experience working closely with high-throughput sequencing technologies. And in addition to my research work, I keep up to date on other researchers' work and the general science in the field by regularly reviewing research published in leading journals in the field.

22. Following my postdoctoral research, I also have over 10 years of experience working with high-throughput sequencing technologies, including next generation sequencing technologies, through my work as a principal investigator overseeing assay development and translational research studies and my role as a lead investigator at the Gerstner Center for Cancer Diagnostics and Blood Biopsy Team at the Broad Institute.

VII. CLAIM CONSTRUCTION

23. I have been informed that claim terms are usually given their ordinary and customary meaning that is based on what a POSA would have understood the term to mean at the time of the invention, but that the meaning must be consistent with how the term is used in the specification. Based on this understanding, I have read the claim language in light of how the terms are used in the intrinsic evidence,

including the specification and prosecution history, and how they would have been understood by a POSA at the time of the invention.

24. I understand “template cross-over error” and “cross-over error” in the ’852 patent to mean “error from sequences of chimeric molecules created during cluster amplification and made from template nucleic acids derived from different samples,” which is a slightly different meaning from that proposed by Petitioner and Dr. Spellman. I disagree with Petitioner and Dr. Spellman’s construction because they do not define a “template cross-over error” in the context of being an error, but rather say a “template cross-over error” is the sequence itself that makes up a chimeric molecule. My understanding is based on the claims and specification, including the problem described and addressed by the ’852 Patent.
25. First, the claimed method explains that it is directed to reducing cross-over errors that result from chimeric molecules. EX1001, 20:65-66 (claim 1). For example, the last limitation of Claim 1 says that cross-over error is reduced by discarding sequence reads obtained from chimeric molecules, not by stopping the formation of the sequences themselves. EX1001, 21:16-19 (claim 1). This is consistent with the description of errors throughout the specification, which describes

“false positives” and “false negatives” as arising from chimeric molecules. EX1001, 7:37-38. I do not read the claims or any of these portions of the specification as requiring reducing the formation of the chimeric sequences themselves.

26. Second, based on the claims, specification, and prosecution history, I understand “cross-over error” to mean the error caused by chimeric sequences that arise after sample pooling and during cluster amplification in the claimed methods. As I also describe in Section VIII below, cross-over error is used in a consistent way in the specification to mean errors, like those formed by “improper molecular overlap,” that occur during the amplification stage after samples have been pooled. EX1001, 4:42-45, 5:1-5, 5:29-28, 5:41-46, Fig. 2. As an example, the specification describes cross-over error as occurring from the formation of heteroduplexes during amplification, and explains how chimeric molecules are formed and further amplified as a result. EX1001, 4:42-45, 5:43-46, 6:35-44; EX2002, 99:15-100:10. The ’852 Patent also explains that, when formed “early in the amplification process,” chimeric sequences can create high-quality signals that cause sequencing data to be incorrectly assigned to a sample. EX1001, 4:60-63.

27. The '852 Patent uses “cross-over error” in the same context (i.e., arising from the formation of chimeric molecules during amplification after sample pooling) when describing the sample identification issues arising from using only one sequence identifier, as was done before the '852 Patent's invention. EX1001, Fig. 1, 4:17-20. For example, the specification says that chimeric molecules are formed after “the amplification process begins” when two template molecules overlap. EX1001, 4:20-23, 4:32-33, 4:40-45, 4:49-52. The resulting chimeric molecules can create a homogenous clonal chimeric cluster on the flow-cell if they are formed early on during the amplification process, multiplying the chimeric sequences. EX1001, 4:57-63. This is consistent with my understanding of how solid-phase amplification occurs on the surface of a flow-cell, because a molecule formed early on during amplification will be exponentially multiplied through bridge-amplification during its many amplification cycles.
28. The '852 Patent then provides a solution to cross-over error from the formation of chimeric molecules during amplification by suggesting the use of unique dual barcodes to accurately identify and discard them during analysis. As I explain at the beginning of this section, the possibility of chimeric molecules being formed does not change just

because two barcodes have been added to a template molecule, because chimeric molecules may still be formed during cluster amplification. EX1001, Fig. 2, 5:1-5, 5:32-33, 5:41-46. The specification explains that the two unique barcodes can reduce cross-over error because identification of one barcode from each sample in a template read will indicate a chimeric molecule needing to be discarded, instead of having only one sequence identifier correlating the chimeric molecule to the wrong sample. EX1001, 5:55-58, 5:65-6:4. The specification also provides Figure 3, which shows the benefits of using dual barcodes instead of single barcodes to reduce incorrect sample identification of sequence reads resulting from chimeric molecules formed during amplification. EX1001, 6:7-9, 6:29-34, 6:41-52, 7:20-23, 7:27-32.

29. I understand that the '852 Patent additionally states that other sequencers in addition to flow-cell based machines can be used to perform its methods and that chimeric sequences might be formed by additional mechanisms. EX1001, 20:37-45, 6:25-26. My understanding of "cross-over error" is consistent with both of these disclosures, which align with the patent's explanation that requires a chimeric molecule to be formed during amplification, as I describe above and throughout Section VIII below. EX1001 Figs. 1-3, 5:1-5, 5:28-29, 5:41-43. For

example, although the specification describes a heteroduplex molecule as one potential way that a chimeric molecule can be formed during amplification, other mechanisms can still occur during amplification to create chimeric molecules. EX1001, 6:25-29. I do not read the claims of the '852 Patent as describing the formation of a "chimeric sequence" at any other point during multiplexed sequencing outside of the cluster amplification step that is performed after sample pooling.

30. My understanding of the meaning of "cross-over error" is also consistent with the prosecution history of the '852 Patent. I understand that the Examiner only allowed the '852 Patent after Applicant amended the claims to recite amplification of template molecules as occurring "on a surface of a flow cell" to form clusters, at least one comprising a chimeric sequence. EX1007, 302, 311. Because an amplification cluster on the surface of a flow cell must contain a chimeric sequence, my understanding and reading of "cross-over error" as occurring during amplification is consistent with Applicant's amendment.

31. I note that Petitioner and Dr. Spellman propose construing all other claim terms based on their ordinary and customary meaning. Pet., 22-23; EX1003 (Spellman Declaration), ¶ 123. For the purposes of this

declaration, I have likewise interpreted the remaining claim terms to have their ordinary and customary meaning.

VIII. TECHNICAL BACKGROUND

32. The '852 Patent and the related U.S. Patent No. 11,768,200 (the "'200 Patent") are entitled "Methods for Maintaining the Integrity and Identification of a Nucleic Acid Template in a Multiplex Sequencing Reaction." The inventors on these patents are Drs. Gregory Porreca, Mark Umbarger, and George Church.
33. The '852 and '200 Patents are about a method for making sample identification more accurate in multiplexing sequencing reactions.
34. The basic steps of multiplex sequencing can be thought of as follows. First, nucleic acids (e.g., DNA) are collected from more than one source (e.g., samples are collected from two patients, A and B). DNA can be collected from any type of biological sample, including blood. EX1001, 7:58-61. For example, blood contains cell-free DNA released from cells that can be isolated using standard DNA purification techniques.
35. Second, identifier sequences (e.g., barcodes) may be attached (e.g., by using a DNA ligase or DNA polymerase) to each nucleic acid molecule (or a copy thereof, if a DNA polymerase is used). These barcodes can be specific and unique to the nucleic acid molecules from each sample

so that each copy can eventually be traced back to its sample source.

Adapters (e.g., P5 and P7 adapters for Illumina sequencers that are required for binding the molecules to the oligonucleotides on the surface of the flow cell) may be added at this stage along with the tags (e.g., through the same primer), or they may be added after sample pooling, as I describe below. Before the '852 Patent, one method of identifying the source of nucleic acid molecules involved attaching only one sample identifier sequence to the 5' or 3' end of each nucleic acid molecule. *See* EX1001, 6:17-20, Fig. 3. The inventions of the '852 and '200 Patents, however, require using two identifier sequences, each unique to the template nucleic acid molecules of each sample and thus used only for that sample, for more accurate sample identification during downstream data analysis. EX1001, 4:53-63, 6:17-20, 7:20-23, Fig. 3.

36. Third, because there may be only a small amount of DNA available in each source, one can optionally choose to perform a first amplification reaction (e.g., 10-20 cycles of PCR) to create some amplicons of the nucleic acid molecules of interest (e.g., a library of tagged nucleic acid molecules with P5 and P7 adapters attached) from that source. If

performed, this amplification reaction occurs prior to the subsequent cluster amplification on a flow cell that I describe below.

37. Fourth, the tagged nucleic acid molecules (e.g., library) from each source are pooled together for further amplification and sequencing of the resulting amplicons. Adapter sequences to facilitate binding of the tagged nucleic acid molecules to the flow cell may be added here, after sample pooling, if they were not added previously. At the time of the '852 Patent, two available sequencing platforms were Illumina sequencers (formerly known as Solexa sequencers) and Roche 454 sequencers. Illumina sequencers were thought to be most useful for sequencing shorter reads, while Roche 454 pyrosequencing was thought to be useful for generating longer reads. EX1005 (Gloor), 2; EX1004 (Parameswaran), 1. Roche 454 pyrosequencing also had several limitations that had to be considered when designing experiments. *See* Section X.A (below).

38. In terms of differences between amplification, Illumina flow-cells involve solid-phase amplification, specifically bridge amplification, where template nucleic acid molecules hybridize to oligonucleotide primers attached to a solid surface via the appended adapter. The primers are extended to create copies, and the original templates are

denatured and removed. Then, each copy bends over to anneal to another nearby complementary primer, forming a bridge, and this nearby complementary primer is extended to create another copy. EX1001, 19:4-16. The resulting copies are denatured, and this process is repeated for many cycles of amplification to form clusters of amplicons attached to the surface of the flow cell that can be sequenced. EX1001, 15:7-12. Roche 454 pyrosequencing, on the hand, uses emulsion PCR, which amplifies nucleic acid fragments on microbeads, and the amplicons are then sequenced. Emulsion PCR is not used with Illumina flow-cells as it involves amplifying nucleic acid fragments in droplets and amplicons accumulate on the surface of a bead (e.g., microbeads). It also does not involve the formation of a bridge structure, nor does it occur on a fixed surface. In both cases, a single template nucleic acid is intended to be clonally amplified to form a cluster. However, at times, more than one template nucleic acid molecule may seed each bridge amplification or emulsion PCR.

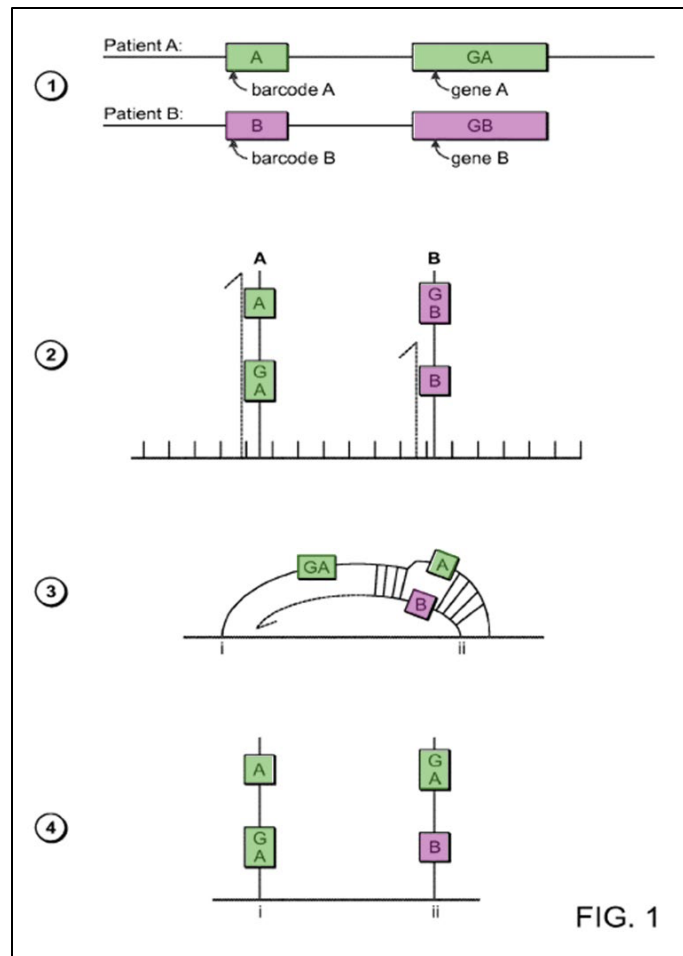
39. Last, the sequencing data of each source can be analyzed by tracing amplicons back to their source using the sample identifier sequences. This analysis allows sequences of interest (e.g., gene fragments) to be appropriately associated with the correct sample.

40. Issues can occur during multiplex sequencing workflows that can make it difficult to trace molecules back to their original sample source. The '852 Patent recognized “improper molecular overlap,” which could arise, for example, from heteroduplex formation, during the amplification performed on flow-cells after sample pooling. EX1001, 4:17-20, 4:42-45, 5:43-46. When “improper molecular overlap” occurs, a chimeric molecule is formed, which may be composed of part of the sequence from a template molecule from one source and part of the sequence from a template molecule from a different source and further amplified. EX1001, 4:42-52, 5:41-48, 6:24-33, claim 1. The formation of these chimeric molecules led to what the '852 Patent calls cross-over error, which it explains can result in “false positive and/or false negative results.” *See, e.g.*, EX1001, 2:8-11, 7:37-38.
41. Before the '852 Patent, in some applications only one sequence identifier was used to tag a nucleic acid molecule, and chimeric molecules could easily go unrecognized, leading to sequences of interest (e.g., gene fragments) being assigned to the wrong sample. EX1001, 4:60-63. Chimeric molecules would form due to a “barcode swap” during amplification, which would make the resulting amplicons (i.e., amplicons resulting from the chimeric molecule) trace to the

incorrect sample (e.g., an amplicon with a gene fragment from Sample B but with a barcode attributable to Sample A). EX1001, 4:17-63, Fig.

1. Without a second barcode attached to the amplicon that was individually uniquely assigned to that amplicon to act as a second check, it could be impossible to tell whether the nucleic acid molecules were being traced to the correct sample. EX1001, 4:53-55.

42. Figure 1 of the '852 Patent (reproduced below) shows the problem of using only one sequence identifier for a template molecule. Panel 3 shows the chimeric molecule containing a barcode associated with Patient B's sample but a gene fragment from Patient A. EX1001, 4:49-52. The sample identification issues resulting from the formation of these chimeric molecules were particularly prominent when the overlap between molecules arose at the beginning of cluster amplification, when several following amplification cycles led to the chimeric molecule being amplified to form many copies and leading to the creation of a high-quality signal. EX1001, 4:55-63.



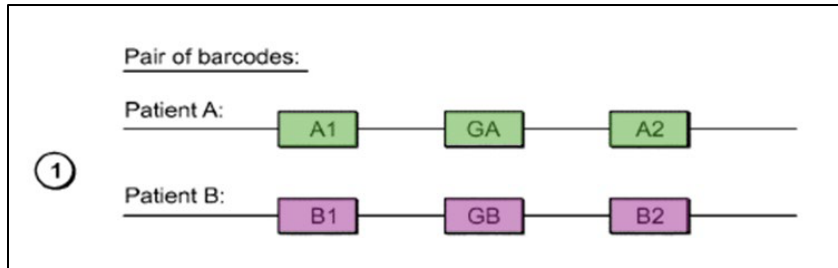
EX1001, Fig. 1 (colors added to show barcode swap when only one sequence identifier is attached to a template molecule).

43. At the time of the '852 Patent, the industry did not appreciate the cross-over error caused by chimeric molecules formed during cluster amplification. For example, many scientists using Illumina sequencers were not aware of the frequency with which chimeric molecules formed, or how they caused cross-over error and skewed sample identification data, until several years later when Illumina publicly

recognized the problem. EX2004 (Kircher 2011); EX2003 (Illumina UDI White Paper).

44. The '852 Patent recognized this problem, however, and disclosed using two unique sample identifiers for all template nucleic acid molecules coming from one sample to improve sample identification in multiplexing. The '852 Patent provided a method for reducing cross-over error that included identifying and discarding amplicons resulting from chimeric molecules formed during amplification on a flow-cell surface after sample pooling. EX1001, 5:1-5, 5:29-28, 5:41-43. The '852 Patent's method describes using "two or more identifiers" with each template molecule so that "the identifiers will only be present when the analyte of interest is detected." EX1001, 2:4-11, 2:16-18. By attaching a unique sequence identifier to both the 5' and 3' portion of a template molecule, a chimeric molecule formed during amplification can be readily identified, by recognizing that one of the identifiers is associated with one sample (e.g., Sample A) while the other is associated with a different sample (e.g., Sample B), and discarded from analysis. EX1001, 2:23-25.
45. Figure 2 of the '852 Patent (reproduced in sections, below) shows how the invention works on a flow-cell. Panel 1 shows samples collected

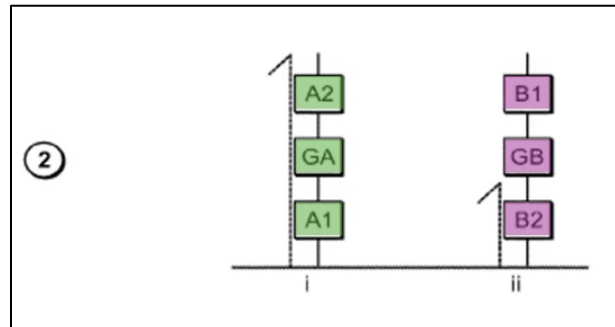
from two different patients, Patient A and Patient B, where sequence identifiers unique to each patient (i.e., A1/A2 for Patient A and B1/B2 for Patient B) have been incorporated on either side of a gene fragment from each patient (i.e., gene fragment A for Patient A and gene fragment B for Patient B). Because each sequence identifier is unique to the nucleic acid molecules from each patient, each of the molecules can be traced back to the patient they came from. EX1001, 5:6-8.



EX1001, Fig. 2, Panel 1 (colors added).

46. Panel 2 of Figure 2 shows the pooling step. Each of the molecules has hybridized to complimentary oligonucleotide primers attached to the surface of the solid support in close enough proximity to each other to set up a potential index hop, which would involve the formation of a chimeric molecule. EX1001, 5:28-40. Panel 2 shows that the primer attached to the template molecule from Patient B has only partially extended, incorporating sequence identifier B2 into the resulting amplicon, but not sequence identifier B1. Following amplification, the template-primer duplex is denatured and each of the resulting

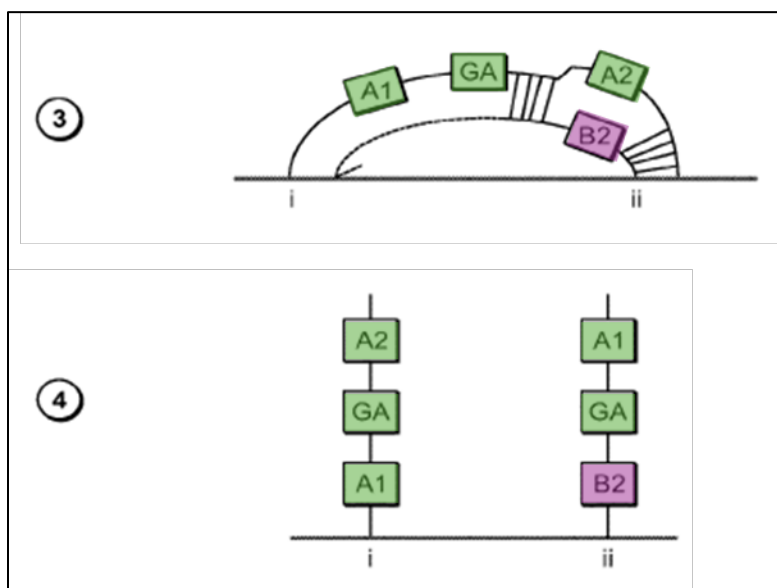
amplicons will re-anneal for subsequent amplification. EX1001, 5:41-43.



EX1001, Fig. 2, Panel 2 (colors added).

47. Panel 3 of Figure 2 (reproduced below) shows that the amplicon molecule (“strand i”) formed from the proximal template from sample A (green) was extended across the full length, and because barcode region A2 may share enough priming sequence surrounding it that is common to either side of barcode region B2, the two amplicons may hybridize with each other to form a bridge, shown in panel 3. EX1001, 5:42-50. This resulting structure is an example of a heteroduplex. EX1001, 5:43-46. Panel 3 also shows strand ii’s extension using a polymerase, which will copy Patient A’s gene fragment and A1 barcode into strand ii containing Patient B’s B2 barcode, forming a chimeric molecule. EX1001, 5:52-55.
48. Then, as shown by Panel 4, because two unique barcode sequences were used instead of one, the resulting chimeric molecule will have a

barcode sequence from both Patient A and Patient B, which will be recognized during sequencing as an invalid barcode pair even if it yields a high-quality signal from the fact that it was created shortly after amplification began. EX1001, 5:62-6:4. The '852 Patent method then excludes any sequences with an invalid barcode pair as chimeras that might lead to cross-over error. EX1001, 6:4-6.



EX1001, Fig. 2, Panels 3 and 4 (colors added).

49. As I describe below, the method disclosed in the '852 Patent was a novel sequencing method to reduce cross-over error in multiplex sequencing reactions, an error that was not appreciated by others at the time of the invention. EX1001, 6:53-57; 7:24-38; Fig. 3; *see also* EX2004 (Kircher 2011); EX2003 (Illumina UDI White Paper).

IX. PROSECUTION HISTORY OF THE '852 PATENT

50. The Applicant applied for the '852 Patent on January 14, 2021. EX1001 (cover). I understand December 23, 2010, to be the effective priority date of the '852 Patent for purposes of the Petition. Pet., 17.
51. I understand the Examiner considered Parameswaran, the same prior art reference now asserted by Illumina in these proceedings, during prosecution and rejected the '852 Patent's then-pending claims as anticipated because of it. EX1007, 195-196.
52. In response, and after an interview with the Examiner, the Applicant amended then-claim 47 to require amplifying nucleic acid templates "on a surface of a flow cell." EX1007, 302. During the interview, the Examiner explained that adding a requirement that amplification arise through mechanisms on the surface of a flow-cell to generate at least one cluster with a chimeric sequence would make the invention distinguishable from the prior art. EX1007, 311. The Applicant in amending the claims explained that the amendment overcame Parameswaran because it required cluster amplification on the "surface of a flow cell" instead of by using other mechanisms, such as emulsion PCR as Roche 454 pyrosequencing used. *See* EX1007, 302, 305. During the interview, the Examiner explained that adding a requirement

that amplification arise through mechanisms on the surface of a flow-cell to generate at least one cluster with a chimeric sequence would make the invention distinguishable from the prior art. EX1007, 311.

53. Following this amendment, the Examiner allowed the claims. EX1007, 312, 316-317.

X. ILLUMINA’S ASSERTED PRIOR ART

A. Parameswaran

54. Parameswaran et al., titled “A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing,” discloses a barcoding approach for high-throughput pyrosequencing of multiple samples. EX1004 (Parameswaran), 1.
55. Parameswaran provides “a *pyrosequencing-tailored* barcoding approach that allows for the unambiguous assignment of nucleic acid sequences ... from up to 48 different samples.” EX1004 (Parameswaran), 8. Parameswaran describes the limitations of multiplexed pyrosequencing, such as physical-space segregation of the sequencing platform, and explains that it “devised a barcoding approach *to substantially enhance* the scope and capacity of multiplexed high throughput *pyrosequencing*.” EX1004 (Parameswaran), 4.

56. Parameswaran uses two unique 10-nucleotide barcodes designed to resolve ambiguities in pyrosequencing. EX1004 (Parameswaran), 2 (“We describe here the design and testing of a series of 10-base barcodes that differ from each other at multiple positions (by at least 4 nt), and have a common specification that facilitates resolution of ambiguities.”). The barcodes are appended to the ends of nucleic acid molecules from each sample, to create barcoded sequencing libraries. EX1004 (Parameswaran), 2-4. Barcoded sequencing libraries from multiple samples are then pooled and sequenced together, and reads are unambiguously assigned to individual samples based on a perfect match to either “barcode motif,” which is the barcode plus additional adjacent nucleotides (e.g., up to 3 nucleotides of a cloning linker). EX1004 (Parameswaran), 1 (“[W]e found that over 99.8% of obtained sequences could be assigned to 25 independent, uniquely barcoded libraries based on the presence of *either a perfect forward or a perfect reverse barcode*.”), 2 (highlighting the ability to “unambiguously assign 99.8% of obtained barcoded sequences by bioinformatically probing for *an error-free barcode at either end of the sequence of interest*.”), 12 (“Sequences that contain both 5’ and 3’ motifs are given first priority, followed by sequences that contain *either the 5’ or the 3’*

motif. Sequences that do not contain perfect matches to any motifs are put into an “unassigned” bin for further analysis.”).

57. According to Parameswaran, its specific design of longer barcodes enabled greater multiplexing and more definitive sample assignment than prior approaches involving much shorter barcodes. EX1004 (Parameswaran), 2 (“Several recent papers have reported the use of very short (2nt- or 4-nt) barcodes with pyrosequencing platforms. Although such approaches are valuable, there are applications that call for *a more definitive assignment of samples and/or enhanced multiplexing capabilities*.”), 8 (“Previously, multiplexed pyrosequencing using the 454 sequencing technology was implemented to analyze sequences from several datasets simultaneously. However, the barcodes used in these experiments are too short to allow for massively parallel runs involving large numbers of sample libraries. Shorter barcodes also have a steeper trade-off between number of possible barcodes and the minimum number of nucleotide differences between individual barcodes.”). Parameswaran describes barcodes that are sufficiently long and distinct to overcome common errors in pyrosequencing such as homopolymeric runs. EX1004 (Parameswaran), 4 (“Length of the barcode was chosen to allow for

definitive assignment of barcodes, while minimizing the number of base reads used per barcode. The barcode design (Table 1) facilitated unambiguous assignment, with care taken to avoid homopolymeric runs that are particularly problematic in pyrosequencing technologies.”).

58. Parameswaran specifically designed its barcodes for use in pyrosequencing technologies. EX1004 (Parameswaran), 4 (“Specifications for the barcodes are based on details of pyrosequencing error characteristics.”). Parameswaran’s barcodes were designed to “reduce errors related to pyrosequencing” and Parameswaran reported that “the experimentally observed spectrum of errors did not match the types of homopolymer-related errors most frequently observed for pyrosequencing.” EX1004 (Parameswaran), 7.
59. Parameswaran discusses three different sequencing systems available in 2007: the Genome Sequencers (GS20 or GS-FLX) from Roche/454 Life Sciences, the 1G Analyzer system from Illumina/Solexa, and the SOLiD system. EX1004 (Parameswaran), 1 (“Three major high-throughput sequencing platforms are in use today: the Genome Sequencers from Roche/454 Life Sciences [GS-20 or GS-FLX;(1)], the 1G analyzer from Illumina/Solexa (2) and the SOLiD System from

Applied Biosystems.”). Parameswaran acknowledges that at the time the Illumina and SOLiD sequencing systems provide many more, shorter reads than the Roche/454 pyrosequencers. EX1004 (Parameswaran), 1 (“Comparison across the three platforms reveals a trade-off between average sequence read length and the number of DNA molecules that are sequenced. At present, the Solexa and SOLiD systems provide many more sequence reads, but render much shorter read lengths than the 454/Roche Genome Sequencers.”). A POSA would recognize that Parameswaran developed a barcoding approach for pyrosequencing microbiomes as opposed to Illumina sequencing systems because, at the time of Parameswaran, Illumina’s sequencing reads were too short to sequence regions of interest. EX1004 (Parameswaran), 1 (“This makes the 454/Roche platform appealing for use with barcoding technology, as the enhanced read length facilitates the unambiguous identification of both complex barcodes and sequences of interest.”); EX1005 (Gloor), 2 (“Until recently, the Illumimna [sic] sequencing-by-synthesis method of parallel DNA sequencing was thought to be unsuitable for microbiome profiling because the sequencing reads were too short to traverse any of the 16S rRNA variable regions.”). A POSA further would understand that in

2007, further advancements to the Illumina sequencers were needed to adapt Parameswaran's barcoding approach to Illumina sequencers. EX1004 (Parameswaran), 1 ("This makes the 454/Roche platform appealing for use with barcoding technology, as the enhanced read length facilitates the unambiguous identification of both complex barcodes and sequences of interest.").

60. Parameswaran speculates that its pyrosequencing-tailored barcodes may be compatible with other platforms, but suggests that the barcodes should be revised to reflect the sequencing error profiles. EX1004 (Parameswaran), 8 ("Our barcode specification has been tailored for use with pyrosequencing technologies, particularly reflecting restrictions imposed on the generation of homopolymeric runs. These barcodes may also be compatible with other sequencing platforms. Optimally, however the precise structure of the barcodes might need to be revised to reflect sequencing error profiles of the various systems.").
61. Parameswaran imposed strict criteria for barcode design, which allowed for designing only 48 barcodes. EX1004 (Parameswaran), 4 ("Sixty-four barcode pairs were thus designed, with any two forward or reverse barcodes differing in at least 4 nt: one doublet + two singlets (Table 2). These serve as strong unique identifiers for each of the

forward and reverse barcode sets. Due to the terminal nucleotide restriction (on position ‘O’), only 48 barcodes may be used with a single 5’–3’ pair of linkers used in cloning.”). Parameswaran’s barcode design only allowed for 48 samples to be pooled and sequenced at a time. Parameswaran hypothesizes that longer barcodes or variations in cloning linkers might expand the barcode design space (EX1004 (Parameswaran), 8), but a POSA at the time of the invention would understand that longer barcodes would not be feasible in Illumina’s sequencers, which were characterized by their short sequence reads. Parameswaran concludes that the advancement that would allow for *increased sample multiplexing capacity* would be using non-unique (combinatorial) dual-indexing. EX1004 (Parameswaran), 7-8.

62. Parameswaran’s pyrosequencing-tailored barcoding approach began with the isolation of RNA and subsequent conversion of RNA to cDNA, with the addition of cloning linkers to both ends. EX1004 (Parameswaran), 2. Parameswaran’s method requires performing two rounds of PCR using 45-46 nucleotide forward and reverse primers with fully unique barcodes comprising “(i) the 454 amplicon adapter (the amplicon sequencing primer annealing site; the ‘F-adapter or R-adapter’), (ii) the barcode (F-barcode or R-barcode) and (iii) the F-

Cloning-Linker or R-Cloning-Linker (the sequences corresponding to the linkers used for construction of the initial nucleic acid library).”

EX1004 (Parameswaran), 2. Parameswaran discloses pooling two sets of 25 libraries and pyrosequencing the pooled libraries on the GS20 platform (454/Roche). EX1004 (Parameswaran), 4.

63. Despite using two unique barcodes for each sample, Parameswaran concluded that only *one error-free barcode motif* is needed for the unambiguous assignment of 99.8% of its pyrosequencing reads. EX1004 (Parameswaran), 2 (“Using the GS20 platform (454/Roche), these barcodes allowed us to simultaneously pyrosequence small RNA libraries from 25 diverse samples for each pilot run, and unambiguously assign 99.8% of obtained barcoded sequences by bioinformatically *probing for an error-free barcode at either end of the sequence of interest.*”), Abstract (“99.8% of obtained sequences could be assigned to 25 independent, unique barcoded libraries based on the presence of *either a perfect forward or perfect reverse barcode.*”).
64. Parameswaran identified a “remarkably low misassignment rate,” reporting that the frequency of mismatched error-free barcodes is as low as 0.00042% on pyrosequencing technology platforms. EX1004 (Parameswaran), Table 3. Based on Parameswaran’s identification of

the “remarkably low misassignment rate” observed on pyrosequencing technology platforms, Parameswaran concludes that only one error-free barcode is necessary for the unambiguous assignment of samples and therefore includes all sequences with one error-free barcode in its analysis. EX1004 (Parameswaran), 8.

65. Parameswaran’s remarkably low false-discovery rate reveals that less than 0.005% of all sequence reads contained mismatched error-free barcodes that unique dual indexes could identify but combinatorial indexes could not. Parameswaran also recognizes that considerably more (n^2 instead of n) samples could be accurately sequenced if barcodes are repeated among samples, instead of using two fully unique barcodes per sample. EX1004 (Parameswaran), 8 (“The low assignment rate also indicates the potential utility of combinations of forward and reverse barcodes to generate up to 48^2 (=2304) combination barcodes for this set of sequences.”), 1 (“With improved read lengths of pyrosequencers, combinations of forward and reverse barcodes may be used to sequence from as many as n^2 independent libraries.”). Therefore, a POSA would understand that Parameswaran suggests combinatorial indexing for further increasing multiplexing capacity.

66. Parameswaran identified three potential causes of the mismatched error-free barcodes that made up its false discoveries: (1) errors in barcode synthesis or sequencing, (2) contamination in PCR primers, or (3) cross-contamination in the second round of PCR. EX1004 (Parameswaran), 7 (“Perfect, ‘unexpected’ occurrences may be attributed 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.”). Parameswaran suggests errors in barcode synthesis and sequencing as likely causes for the mismatched error-free barcodes. EX1004 (Parameswaran), 7 (“[T]he observed error spectrum *appeared most closely to match that expected as a result of known imprecisions in the chemical synthesis of DNA primers* used in standard PCR reactions.”); Table 3 (“The number of sequences with imperfect motifs also correspondingly increased with the increased length of the barcode motif, *reflective of the high error rate of the pyrosequencer technology*.”). Parameswaran does not identify cross-over error as an issue for its dataset, and instead considered “all sequences with a match to one of the two expected

sequence motifs ... as true members of the corresponding dataset.”

EX1004 (Parameswaran), 8.

67. Dr. Spellman states that “Parameswaran’s unique dual-indexing method permitted ‘unambiguous assignment’.” EX1003 (Spellman Declaration), ¶ 161. Specifically, Dr. Spellman states “[t]o facilitate ‘unambiguous assignment of nucleic acid sequences from a mixture of libraries from up to 48 different samples,’ Parameswaran discloses ‘requiring a match to the 10-base barcodes and additional three bases of the fixed linker’—called a ‘barcode motif.’” EX1003 (Spellman Declaration), ¶ 126. Dr. Spellman misreads Parameswaran.
68. Parameswaran did not use dual indexes to unambiguously assign sequence reads to individual samples. Instead, Parameswaran discovered that it could “unambiguously assign 99.8% of obtained barcoded sequences by bioinformatically probing for an error-free barcode *at either end* of the sequence of interest.” EX1004 (Parameswaran), 2; *see also* EX1004 (Parameswaran), 1 (“[W]e found that over 99.8% of obtained sequences could be assigned to 25 independent, uniquely barcoded libraries based on the presence of *either a perfect forward or a perfect reverse barcode.*”). Indeed, Parameswaran’s supplementary information confirms that

Parameswaran only requires one perfect barcode per sequence to unambiguously assign 99.8% of obtained sequences. EX1004 (Parameswaran), 13 (“99.87% of sequences in Run I and 99.83% of sequences in Run II had at least *one barcode motif* (10 nt. barcode + 3 nt adjoining cloning-linker) with no errors.”). Thus, the sequencing and filtering of a single barcode motif permitted the unambiguous assignment of samples in Parameswaran.

69. Dr. Spellman also states that “Parameswaran’s unique dual-indexing method permitted ... a ‘remarkably low misassignment rate.’” EX1003 (Spellman Declaration), ¶ 161. But Parameswaran merely identified a low misassignment rate on pyrosequencing platforms. Because of this low misassignment rate, Parameswaran suggests using combinatorial indexing as opposed to unique dual indexing. EX1004 (Parameswaran), 8 (“The low misassignment rate also indicates the potential utility of combinations of forward and reverse barcodes to generate up to 48^2 (=2304) combination barcodes for this set of sequences.”), 1 (“With improved read lengths of pyrosequencers, combinations of forward and reverse barcodes may be used to sequence from as many as n^2 independent libraries for each set of ‘n’ forward and ‘n’ reverse barcodes,”). Indeed, Parameswaran’s comment, “[t]he low false-

discovery rate 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) supports the conclusion that combinatorial tagging (which requires using both 5' and 3' barcode motifs) should be feasible. A POSA would understand that Parameswaran's identification of a remarkably low misassignment rate permitted the use of combinatorial indexes, as opposed to unique dual indexes, in future experiments. EX1004 (Parameswaran), 8.

B. Gloor

70. Gloor et al., titled "Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products," discloses a method of combinatorial indexing for multiplexed microbiome profiling of many samples at low cost using Illumina sequencing. EX1005 (Gloor), 1.
71. Gloor reports on the feasibility of sequencing many microbiome samples together for microbiome profiling using Illumina sequencers. EX1005 (Gloor), 1. Gloor acknowledges that pyrosequencing originally provided an improvement for microbiome profiling over previous methods, such as "Sanger sequencing," but explains that pyrosequencing had limitations. EX1005 (Gloor), 1. Gloor further

explains that “[u]ntil recently, Illumimna [sic] sequencing-by-synthesis method of parallel DNA sequencing was thought to be unsuitable for microbiome profiling because the sequencing reads were too short to traverse any of the 16S rRNA variable regions.” EX1005 (Gloor), 2.

72. Gloor explains that Illumina sequencing is “an approximate order of magnitude cheaper than the per-read cost of pyrosequencing.” EX1005 (Gloor), 2. Gloor further observes that Illumina sequencing platforms allow for “greater depth at a much lower cost” than the current pyrosequencing technology. EX1005 (Gloor), 23. Because of those advantages, Gloor specifically designed its method for Illumina sequencers.

73. Gloor describes the PCR amplification of the V6 region of the 16S rRNA gene of each sample using a primer pair with a distinct combination of sequence tags. EX1005 (Gloor), 3. Gloor’s PCR products with combinatorial sequence tags were pooled with PCR products from other samples and sequenced. EX1005 (Gloor), 1. Gloor assigns 76-nucleotide “paired end” reads from PCR products with perfect overlapping region matches and two valid sequence tags to the correct samples. EX1005 (Gloor), 7. By using a combination of sequence tags, Gloor highlights that one advantage over other

approaches is the ability to sequence many samples with a small number of primers. EX1005 (Gloor), 1 (“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.”).

74. Gloor uses “combinatorial sequence tagging ... to examine hundreds of samples with far fewer primers than is required when sequence tags are incorporated at only a single end.” EX1005 (Gloor), 1. For example, Gloor used 18 left sided sequence tags and 16 right sided sequence tags to sequence 272 clinical samples in a single Illumina run. EX1005 (Gloor), Table 2. Combinatorial sequence tagging creates unique pairs for each sample, by adding non-unique, or repeat, forward and reverse tags to the template nucleic acids from each patient. Using Gloor’s tags, a POSA would understand that there are $(18 \times 16) = 288$ potential combinatorial indexes but only 16 potential unique dual indexes. Thus, a POSA would understand that combinatorial tagging makes sequencing more efficient and less costly, as it allows for more samples to be sequenced with fewer unique barcodes. Because combinatorial sequences use the same tags in different combinations for different samples, chimeric molecules that form during the amplification step, with one tag from Sample A and another tag from Sample B, would be

incorrectly identified and analyzed as sequences from a different sample. Gloor did not acknowledge cross-over error in any way, and a POSA at the time would not have recognized that cross-over error was a problem in Gloor.

75. Gloor designed sequence tags with varied lengths from 3 to 6 nucleotides to ensure all 4 nucleotides were present in the first 4 positions of each primer. EX1005 (Gloor), 5. Gloor explains this variation helps ensure that adjacent clusters are not mistaken for one another. EX1005 (Gloor), 5. A POSA would have therefore understood that varying the lengths of designed sequence tags would limit the likelihood of adjacent clusters being scored as one. EX1005 (Gloor), 5 (“The tag length variation was expected to reduce the likelihood that adjacent spots on the Illumina solid support would be scored as one during the sequencing of the amplification primers.”).
76. Gloor applied its method to analyze 272 clinical samples of the vaginal microbiota from HIV+ women. Gloor obtained 18,047,860 read pairs of 76 nucleotides from both ends of PCR products, of which 12,035,329 read pairs had “two valid sequence tags and both primer sequences” (allowing up to 3 mismatches per primer sequence) and perfect overlaps. EX1005 (Gloor), 7. Only reads with perfect overlaps (e.g., 0

mismatches in the overlapping region) were used for analysis. EX1005 (Gloor), 7. Gloor grouped the 12,035,329 read pairs with two valid sequence tags, both primer sequences (allowing up to 3 mismatches per primer), and 0 mismatches in the overlap region, into 126,832 identical sequence units (ISUs) which shared the same sequences between the primers. EX1005 (Gloor), 8.

77. Gloor identified multiple sources of error, none of which were cross-over error. First, Gloor found “sequences with incorrect sequence tags,” which included “tags” that did not match any of the predetermined tags introduced during the library preparation step. Gloor presumed that those incorrect sequences were largely due to one-nucleotide truncations, which they speculated had arisen during primer synthesis. EX1005 (Gloor), 4 (“Inspection of sequences with incorrect sequence tags showed that the single largest contributor to the difference between the number of reads with proper primer sequences and the number of reads with proper sequence tag sequences was an N-1 truncation of the sequence tag, which presumably arose during the primer synthesis.”).
78. Second, Gloor found “reads missing sequence tags,” which a POSA would understand to be unusable under any barcoding technique. EX1005 (Gloor), 4 (“The next largest class of sequence tag error was

complete lack of the left or right end sequence tag.”). A POSA would understand that these reads would be unusable for dual-indexed sequencing, regardless of whether libraries are uniquely or combinatorially tagged, if both tags are required for read assignment.

79. Third, Gloor attributed the remaining sequence tag errors to a variety of errors speculated to arise during the first PCR amplification of each sample, which, as explained above occurs before sample pooling or sample sequencing. EX1005 (Gloor), 4 (“The remaining missing reads are composed of a large number of classes of sequence tag sequences each containing small numbers of errors including additional 5’ bases, misincorporated bases or difficult to classify errors that presumably arose during the PCR amplification.”). A POSA would understand those errors to be different than those arising from any cross-over event that could occur between nucleic acid templates from two different samples.

80. Gloor observed “very few chimeric sequences” in its dataset. EX1005 (Gloor), 22. Because chimeras were known to arise during PCR amplification of individual samples, as opposed to the amplification that occurs after pooling, or due to overlapping clusters on the flow cell without recombination (which are distinct from crossover errors),

Gloor applied the UCHIME algorithm to identify incorrect sequences formed by the recombination of other sequences. EX1005 (Gloor), 4-5. Gloor identified “5211 putative chimeric ISUs” amounting to 5.6% of all ISUs and 0.15% of all reads. Examination of the most abundant putative chimeric ISUs informed refined UCHIME thresholds which yielded only 497 chimeric ISUs from 0.015% of all reads. EX1005 (Gloor), 8. Because “chimeric sequences composed a very small subset of the total number of ISUs,” Gloor concluded that chimeric sequences were negligible in their dataset and included such sequences in its analysis “without further regard.” EX1005 (Gloor), 5 (“The dataset was used without further regard to chimeric sequences because putative chimeric sequences composed a miniscule fraction of reads.”).

81. Gloor speculated that its method yielded “very few chimeric sequences,” because it “used a relatively small amount of input DNA and used a small number of PCR cycles for amplification.” EX1005 (Gloor), 22. A POSA would have also understood such methods to reduce chimeric sequences of the type Gloor discusses, as the PCR amplification prior to sample pooling of each individual sample was known to cause chimeric sequences *within individual samples*. A POSA would therefore have understood the Gloor chimeric sequences

to be made from the nucleic acids within each sample, prior to sample pooling.

82. Gloor reasoned that because the primers were only amplified across a variable region, the “chimeric sequences composed a very small subset of the total ISUs.” EX1005 (Gloor), 8. In other words, Gloor suggests that one may expect less recombination to occur between DNA fragments within each sample when the region that is being amplified within each sample varies in sequence among fragments. Gloor further speculated that many of the chimeric sequences were removed by requiring proper sequence tags and overlapping segments of defined length. EX1005 (Gloor), 22 (“[M]any chimeric sequences may have been removed because of the strict requirement for proper sequence tag and primer sequences on the left and right ends, and because of a requirement for long overlapping segments of a defined length.”).
83. Gloor also hypothesized that requiring an overlap in read pairs helped limit chimeric sequences. EX1005 (Gloor), 22 (“In this case, the deterministic read lengths of the Illumina protocol combined with our narrow window for overlapping segments would have resulted in many chimeric sequences being filtered out.”), 3 (“the V6 region was expected to produce amplified fragments between 110 and 130 bp,

ensuring that the majority of paired-end reads would overlap.”), 4 (finding reads with an “overlap between 10 and 50 nucleotides”). Gloor verified that requiring proper sequences and primers as well as perfect overlapping reads removed chimeric sequences. EX1005 (Gloor), 22 (“[I]nspection of a fraction of the read pairs that failed to overlap, or that failed to pass the sequence tag and primer requirements showed that many of these were chimeric or deleted at one or both ends (data not shown).”).

84. Gloor concluded that Illumina sequencing was highly accurate and “permitted saturating or near-saturating analysis of samples of the vaginal microbiome.” EX1005 (Gloor), 1. Gloor found that the majority of remaining error in its “exceedingly low error rate” was introduced in the PCR amplification of each sample—before samples were pooled and sequenced. EX1005 (Gloor), 2 (“Importantly, we found that the Illumina sequencing method has an exceedingly low error rate and that the majority of errors arise during the PCR amplification step.”).

C. Bentley

85. Bentley et al., titled “Accurate whole human genome sequencing using reversible terminator chemistry,” discloses a method for whole-genome sequencing a single human sample using massively parallel sequencing

by synthesis with the Illumina Genome Analyzer (GA1) sequencer.

EX1006 (Bentley), 1.

86. Bentley explains that it took researchers several years, around \$300 million, and several hundred capillary sequencers to sequence the entire human genome sequence in 2004. EX1006 (Bentley), 1 (“The raw data for the three billion base (3gigabase (Gb)) human genome sequence, completed in 2004, was generated over several years for ~\$300 million using several hundred capillary sequencers.”). Bentley explains that “[m]ore recently an individual human genome sequence has been determined for ~\$10 million by capillary sequencing.” EX1006 (Bentley), 1. Furthermore, Bentley acknowledges that a human genome sequence was recently determined using massively parallel DNA sequencing but notes “much bigger improvements are necessary to enable routine whole human genome sequencing.” EX1006 (Bentley), 1 (“Several new approaches at varying stages of development aim to increase sequencing throughput and reduce cost. They increase parallelization markedly by imaging many DNA molecules simultaneously. One instrument run produces typically thousands or millions of sequences that are shorter than capillary reads. Another human genome was recently determined using one of these approaches.

However, much bigger improvements are necessary to enable routine whole human genome sequencing in genetic research.”). A POSA would understand that Bentley’s goal was to develop a massively parallel sequencing approach that allows for whole human genome sequencing to be feasibly used as a clinical tool. EX1006 (Bentley), 6 (“Massively parallel sequencing technology makes it feasible to consider whole human genome sequencing as a clinical tool in the near future.”)

87. Bentley explains that it sequenced the whole genome of a male human over a period of 8 weeks using six different Illumina Genome Analyzer sequencing platforms. EX1006 (Bentley), 3 (“We sequenced the genome of a male Yoruba from Ibadan, Nigeria.... We generated 135Gb of sequence (~4 billion 35-base reads; see Supplementary Table 6) over a period of 8 weeks (December 2007 to January 2008) on six GA1 instruments averaging 3.3Gb per production run (see Supplementary Table 1 for example). The approximate consumables cost (based on full price of reagents) was \$250,000.”). Bentley sequenced DNA fragments from one human sample at a time and did not use any sample tag, barcode, or identifier. EX1006 (Bentley), 1 (“We demonstrate utility by re-sequencing an individual human

genome to high accuracy.”). Therefore, a POSA would understand that Bentley did not and could not discard or exclude sequences based on identifiers.

88. Bentley describes “mixed clusters that overlap their nearest neighbors” during sequencing and provides strategies to mitigate these. EX1006 (Bentley), 2 (“We discarded all reads from mixed clusters and used the remaining ‘purity filtered’ reads for analysis.”). Of note, Bentley does not speculate that mixed clusters involve recombination of DNA molecules. Indeed, a POSA would understand mixed clusters merely as clusters that overlapped. A POSA would also understand that mixed clusters are affected by the loading density of library molecules onto the flow cell, which can be adjusted. EX1006 (Bentley), 18 (“Depending on the loading density, we typically kept between 50% and 70% of the raw reads.”).

89. Bentley used three separate experiments to show that “purity filtered” reads enabled accurate genetic analysis in multiple datasets. *First*, Bentley sequenced a human bacterial artificial chromosome (BAC) clone. EX1006 (Bentley), 2 (“[w]e used 5 Mb of 35-base purity filtered reads (30-fold average input depth of the BAC) and obtained 99.96% coverage of the reference. There was one consensus miscall, at a

position of very low coverage (just above our cutoff threshold), yielding an overall consensus accuracy of 99.999%.”). **Second**, Bentley sequenced “flow-sorted X chromosomes of a Caucasian female” and found that the purity filtered reads covered 99.8% of the genotyped positions. EX1006 (Bentley), 2 (“We generated 278-million paired 30–35-bp purity filtered reads and aligned them to the human genome reference sequence.... The sequence data covered 99.8% of the 13,604 genotyped positions and we found excellent agreement between sequence-based SNP calls and genotyping data (99.52% or 99.99% using ELAND or MAQ, respectively; Supplementary Table 3).”). **Third**, Bentley sequenced the genome of a man from Nigeria and found that purity filtered reads had a 99.5% concordance with genotyping calls. EX1006 (Bentley), 3 (“We found excellent agreement of our SNP calls with genotyping results: sequence-based SNP calls covered almost all of the 552,710 loci of HM550, with 99.5% concordance of sequencing versus genotyping calls (Table 1 and Supplementary Table 7a).”). A POSA would therefore understand that Bentley’s purity filters successfully mitigated mixed clusters caused by overlapping clusters on the sequencing platform.

**XI. GROUND I: THE CHALLENGED CLAIMS ARE NOT OBVIOUS
OVER PARAMESWARAN AND GLOOR**

**A. A POSA Would Not Be Motivated To Combine Parameswaran
and Gloor.**

90. Dr. Spellman states that a POSA would use Parameswaran’s unique dual indexes with Illumina sequencing platforms as used in Gloor to arrive at the claimed method. EX1003 (Spellman Declaration), ¶ 156. For the reasons below, Dr. Spellman fails to show that a POSA would have been motivated to combine Parameswaran with Gloor or that a POSA would have done so with a reasonable expectation of success of making the claimed method.

**1. Dr. Spellman Does Not Consider The Full
Context Of The References And The State Of The
Art.**

91. Parameswaran is a 2007 article aimed at overcoming limits “in complexity (number of samples sequenced in parallel), and in capacity (number of sequences obtained per sample)” of multiplexed high-throughput pyrosequencing. EX1004 (Parameswaran), 1. Because Parameswaran noted that previous pyrosequencing techniques were limited to 16 independent samples that were physically separated on the sequencing medium, Parameswaran’s goal was to design a method that allowed sequences to be pooled together and sequenced without

physical separators to “maximize[] sequence space and multiplexing capabilities.” EX1004 (Parameswaran), 2 (“The presence of these assigned barcodes allow for independent samples to be pooled together for sequencing, with subsequent bioinformatic segregation of the pyrosequencing output. By not relying on physical separators, this procedure maximizes sequence space and multiplexing capabilities.”); EX2002 (Spellman Tr.) at 147:22-148:3 (“The Parameswaran paper was interested in thinking about how they might move toward very large numbers of samples and coming up with a scheme to handle that”). Parameswaran designed barcodes specifically tailored to the pyrosequencing platform and explained that its provided barcoding method “allows for the unambiguous assignment of nucleic acid sequences from a mixture of libraries from up to 48 different samples.” EX1004 (Parameswaran), 8. Parameswaran further highlighted the potential to assign even more samples with combinatorial tagging. EX1004 (Parameswaran), 8 (“The low misassignment rate also indicates the potential utility of combinations of forward and reverse barcodes to generate up to 48^2 (=2304) combination barcodes for this set of sequences.”) EX2002 (Spellman Tr.) at 147:22-148:3 (“[Q.] The scheme that [Parameswaran] suggested could be used [to move toward

very large numbers of samples] was nonunique dual indexing, true? A.

Yes, I believe that's correct.”).

92. By the time of the '852 Patent in 2010, a POSA would recognize that Illumina sequencers enhanced sequencing capacity beyond what Parameswaran's method was capable of sequencing. For example, Illumina sequencers were known to have “20-fold greater sequence coverage” than pyrosequencing techniques and generated “~12 million usable reads.” EX1005 (Gloor), 2, 23. Indeed, Gloor concluded that it was able to uncover the full sequence diversity of the microbiome in each sample and that deeper sequencing did not identify any new sequence species in its analysis. EX1005 (Gloor), 23 (“The utility of the method is further demonstrated by the near-saturating number of ISU and OTU sequences obtained from a large number of clinical samples. We used several lines of evidence to show that 20000 reads are sufficient to capture all or virtually all of the sequence diversity in the vaginal microbiome, and that obtaining over 50000 reads results in no new sequence species.”). A POSA would understand Gloor's method was sufficiently accurate for the intended use and provided a method to sequence a large number of samples at the required depth.

93. Furthermore, a POSA would understand that Illumina sequencers could pool and sequence hundreds of samples together at a higher complexity and capacity with a lower error rate than Parameswaran's pyrosequencing approach. EX1005 (Gloor), 22. Gloor recognized that using combinatorial tags with Illumina sequencers provided "much greater depth at a much lower cost than is possible with current pyrosequencing technology." EX1005 (Gloor), 23. By 2010, a POSA would therefore recognize that multiplex sequencing on Illumina sequencers had moved beyond the sequencing technology and approach addressed in Parameswaran, and would not have seen any utility in Parameswaran's pyrosequencing-tailored methods on Illumina sequencers.

94. A POSA in 2010 would not look to pyrosequencing technology to improve Illumina sequencers. Gloor cautioned that it is often "unsuitable" to use pyrosequencing methods to analyze data with Illumina sequencing reads, explaining that "[t]he number of and the short length of the reads prevented the application of many common pyrosequencing data analysis pipelines." EX1005 (Gloor), 2. Gloor also provided as context issues with pyrosequencing technologies that led to more error than Illumina sequencing. EX1005 (Gloor), 1. For

example, Gloor acknowledged that using pyrosequencing technologies for microbiome profiling led to rare errors that are not present on Illumina sequencers. EX1005 (Gloor), 1. Gloor explains that because it found pyrosequencing methods were unsuitable for Illumina sequencers, it created its own workflow, including a new barcoding method, “that can be used for rapid and robust generation of the relative abundance of organisms in each sample.” EX1005 (Gloor), 2. A POSA would therefore not look to introduce Parameswaran’s pyrosequencing method into Gloor’s “rapid and robust” workflow. EX1005 (Gloor), 2.

95. Neither Parameswaran nor Gloor suggests that UDI would improve the multiplex capabilities of the Illumina sequencers in Gloor. Instead, Parameswaran suggests the opposite, recognizing that the using combinatorial indexing would greatly increase the number of samples that could be pooled and sequenced. EX1004 (Parameswaran), 4 (“The low misassignment rate also indicates the potential utility of combinations of forward and reverse barcodes to generate up to 48^2 (=2304) combination barcodes for this set of sequences.”). Gloor too recognized that using combinatorial tagging could “examine hundreds of samples with far fewer primers than is required when sequence tags

are incorporated at only a single end” with an “exceedingly low error rate.” EX1005 (Gloor), 22, 2.

96. Gloor was able to sequence 272 samples together using only 34 barcodes. EX1005 (Gloor), Table 2 (showing 18 left sided sequence tags and 16 right sided sequence tags, for *a total of 34 tagged primers*). A POSA would understand that using UDIs in Gloor, instead of the CDIs Gloor used, would require *a total of 544 tagged primers* to sequence all 272 samples in a single Illumina run. This is because for each sample, two unique barcoded primers are needed, so $(272 \times 2) = 544$ primers would be required. A POSA would thus understand that creating such primers and running them would be significantly less efficient and more costly than the CDI method in Gloor.
97. Because Parameswaran identified a low misassignment rate and the utility of combinatorial tagging, and Gloor used CDI and confirmed an exceedingly low error rate, cost and efficiency would have led a POSA away from using far more tagged primers to sequence fewer samples.

**2. Dr. Spellman's Reasons To Combine
Parameswaran and Gloor Fail To Show A Motivation
To Combine Unique Dual Indexes With Illumina
Sequencers.**

**i. Dr. Spellman Does Not Show That Efficiency
And Cost Considerations Would Have
Motivated The Proposed Combination.**

98. Dr. Spellman states that a POSA would have been motivated to combine Parameswaran and Gloor for cost and efficiency reasons. EX1003 (Spellman Declaration), ¶¶ 156-157. Dr. Spellman highlights “the greater efficiency and lower cost of Illumina sequencing” as support that a POSA would be motivated to combine Parameswaran and Gloor. EX1003 (Spellman Declaration), ¶ 157. But Dr. Spellman only focuses on the cost of the sequencers themselves and ignores extra costs associated with Parameswaran’s unique dual indexes. EX1003 (Spellman Declaration), ¶¶ 156-157.
99. Gloor touts a “low-cost, high throughput” method that uses combinatorial sequence tags. EX1005 (Gloor), 1. Gloor explains that Illumina sequencers provide a greater depth than pyrosequencing platforms. EX1005 (Gloor), 23. Gloor highlights the high cost of pyrosequencing technologies as one of the reasons it chose Illumina sequencers. EX1005 (Gloor), 2. Gloor explained that combinatorial tagging helped reduce the number of primers used and a POSA would

understand that a reduction in the number of primers used would lead to a lower cost. EX1005 (Gloor), 22.

100. A POSA would understand that combinatorial sequence tagging provides a cheaper and more efficient method to multiplex sequencing than unique dual indexing. As mentioned above, combinatorial tagging would be less costly than unique dual indexing given the substantially fewer primers required. A POSA would understand that unique dual indexing requires exponentially more unique barcodes to assign a given number of samples than combinatorial indexes. EX1004 (Parameswaran), 8 (noting 48^2 samples could be assigned using combinations of the tags used to assign 48 samples with unique dual indexes). Indeed, if the 272 samples that Gloor were sequenced with unique dual indexes, instead of the combinatorial sequence tags Gloor actually used, a POSA would understand that it would require a total of 544 unique sequence tags, 16 times more than the 34 unique sequence tags (18 left tags and 16 right tags) that Gloor actually uses. EX1005 (Gloor), 6. A POSA would therefore understand that using unique dual indexing would increase the cost of Gloor's low-cost method.
101. A POSA would also understand that unique dual indexes could lead to more observed errors from complications in barcode design, selection,

and identification from sequencing reads. For example, Parameswaran noted that “[s]ince six of the forward barcodes are exact reverse complements of six reverse barcodes, we added a number of adjoining nucleotides in the Cloning-Linker (0-17) to the barcode query to reduce ambiguity while segregating the sequences.” EX1004 (Parameswaran), 7. As shown in Supplementary Table 1A, Parameswaran’s barcodes for samples 43 and 47, samples 45 and 50, and samples 61 and 63 all contain reverse complements of one another. EX1004 (Parameswaran), 14-15. Indeed, the data reported in Supplementary Table 4 of Parameswaran shows that mismatched error-free barcodes were most prevalent between samples 43 and 47, samples 45 and 50, and samples 61 and 63, when only the barcode (and 0 nt of the cloning linker) was used as the search motif. EX1004 (Parameswaran), 20. In Run 1, Parameswaran reported the highest occurrence of mismatched error-free barcodes between samples 4 and 47, samples 45 and 50, and samples 61 and 63, when only the barcode (and 0 nt of the cloning linker) was used as the search motif. EX1004 (Parameswaran), 20. Therefore, a POSA would understand that designing barcodes that are too similar to one another or exact reverse complements from another barcode could lead to more observed errors. Because more unique

barcodes are required when using unique dual indexes, a POSA would further understand that the challenges with designing and creating barcodes would be greater with unique dual indexes than combinatorial tagging. Indeed, Parameswaran suggested “multiple fortuitous errors during barcode synthesis or sequencing” may have resulted in mismatched barcode motifs. EX1004 (Parameswaran), 7.

ii. Dr. Spellman Does Not Show That A Desire To “Reduce Sequencing Error” Would Have Motivated The Proposed Combination.

a. Parameswaran And Gloor Both Claimed A Lack Of Sequencing Errors.

102. Parameswaran used a pyrosequencing approach to minimize pyrosequencing-related errors and reported a “remarkably low misassignment rate (0.000042% and 0.0044% for two independent experiments).” EX1004 (Parameswaran), 8. For example Parameswaran explained that it “expected to minimize errors related to pyrosequencing” and reported that errors did not match the types of errors most frequently observed with pyrosequencing but instead reported that “the observed error spectrum appeared most closely to match that expected as a result of known imprecision in the chemical synthesis of DNA primers used in standard PCR reactions.” EX1004 (Parameswaran), 7. A POSA would have understood that

Parameswaran suggests that its methods had already addressed any errors that arose from the pyrosequencing methods that it used.

103. Gloor reported “that the majority of errors arise during the PCR amplification step.” EX1005 (Gloor), 2. Unlike Parameswaran, Gloor does not have the second round of PCR and performed fewer PCR amplification cycles than Parameswaran. EX1005 (Gloor), 22–24; EX1004 (Parameswaran), 3. Parameswaran suggests that the error-free mismatched barcode motifs arose from the second round of PCR. EX1004 (Parameswaran), 7 (“Perfect, ‘unexpected’ occurrences may be attributed to ... contamination in PCR primer stocks, or to cross-contamination of template DNAs in the second round of PCR.”).
104. A POSA would understand that the second round of PCR is (1) not necessary for multiplex sequencing and (2) creates increased risk for contamination from additional sample handling. For example, Maughan (2009) used a dual barcode strategy in a pyrosequencing experiment with a single round of PCR to ensure that a single barcode was sequenced given the read length restrictions. EX2009 (Maughan 2009), 2. Because Maughan only used a single round of PCR in its pyrosequencing, the additional handling and processing needed to perform a second round was minimized. EX2009 (Maughan 2009), 3.

Maughan shows that when only one round of PCR is performed using primers with dual index tags, there were no mismatched barcodes identified. EX2009 (Maughan 2009), 5 (“Indeed, on average only 13.6% of the reads evidence MID sequences on both sides of the read and, *as expected, in all of these cases the same MID was present on both sides of the DNA fragment.*”). Thus, a POSA would understand that the second round of PCR used in pyrosequencing experiments creates increased risk for contamination in those experiments (e.g., from handling or processing errors such as additional pipetting required).

b. A POSA Would Not Be Motivated To Discard Chimeric Sequences And Thus Reduce Cross-Over Error

105. Dr. Spellman states that “[a]s of December 2010, a POSA would have known that chimeric sequences, *including those resulting from cross-over errors*, were problematic for next generation sequencing that used PCR amplification, including Illumina sequencing.” EX1003 (Spellman Declaration), ¶ 160. But Dr. Spellman fails to show that *cross-over errors* were known or problematic for Illumina sequencing. Indeed, Dr. Spellman provides examples of chimeric sequences which could arise in PCR amplification of a single sample but fails to show

how a POSA would know those chimeric sequences to result from Dr. Spellman's own definition of cross-over error. EX1003 (Spellman Declaration), ¶¶ 160-161.

106. Dr. Spellman asserts that it was known, based on earlier publications that “[p]roblems 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*.” EX1003 (Spellman Declaration), ¶ 67. To support this statement, Dr. Spellman provides excerpts from Lahr, Meyerhans, Paabo, and Odelberg. EX1003 (Spellman Declaration), ¶ 67. Yet, none of these show that PCR recombination of tagged nucleic acids *from different samples* were known to arise *and produce chimeras with cross-overs of sequence tags*.

107. Dr. Spellman states that Lahr shows “[t]he 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)” (EX1003 (Spellman Declaration), ¶ 68), and that “[a]s 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.” EX1003 (Spellman Declaration), ¶ 69. Dr. Spellman also states that “[a] POSA would have understood that the sequences with mismatched barcode motifs in *Parameswaran* included those resulting from template cross-over error introduced during sequencing workflow (sequences of chimeric molecules made from template nucleic acids derived from different samples). EX1017 (Lahr), 1” EX1003 (Spellman Declaration), ¶ 130. But Lahr does not disclose cross-over error or the chimeric sequences with incorrect or mismatched tags that would cause such error. Instead, Lahr describes the “percentages of chimeras formed per reaction” when PCR is performed under varied conditions and concludes that “[c]apturing the full diversity *within a sample* requires a combination of multiple PCR reactions that have been performed under chimera-reducing conditions.” EX1017 (Lahr), 5. Template nucleic acids in Lahr are not tagged, and the “worrisome aspect of PCR-based studies” to which Lahr refers is the chimera formation affecting the determination of the “actual genetic diversity *of a sample*.” EX1017 (Lahr), 1. A POSA would therefore understand Lahr to reference

chimeras within a given sample, not between two different samples, and would understand that UDI methods would have no ability over CDI methods to identify and reduce those chimeras.

108. Meyerhans describes the “PCR co-amplification of two distinct HIV1 *tat* gene sequences” and shows that “PCR mediated recombination may be a problem when studying heterogeneous genetic material such as RNA viruses, multigene families, or repetitive sequences.” EX1018 (Meyerhans), 1. Meyerhans speculates three causes for PCR mediated recombination: Taq DNA polymerase could be pausing on the template; it could be due to premature termination; or it could be preferred sites of nucleation between complementary DNA strands. EX1018 (Meyerhans), 3-4. Meyerhans concludes that “[s]uch recombinant sequences are not expected to be problematic if lower concentrations of template and fewer cycles are used.” EX1018 (Meyerhans), 4. In other words, Meyerhans provides guidance on how to limit the formation of chimeric sequences when PCR is performed. Indeed, Meyerhans does not use tagged nucleic acids and does not discuss PCR recombination between samples, and a POSA would understand that UDIs would not have any impact on those chimeras.

109. Paabo designs a reaction that will only amplify recombination events by PCR and uses this to study the conditions that lead to recombination. Paabo examines PCR recombination induced by breaks in the template, by insertion of adenosine residues such as by *Taq* polymerase, and by DNA damage including in ancient DNA samples. Paabo does not utilize tagged nucleic acids but concludes “when the thermostable *Thermus aquaticus* DNA polymerase encounters the end of a template molecule, it sometimes inserts an adenosine residue; the prematurely terminated product then jumps to another template and polymerization continues, creating an in vitro recombination product. Consequently, amplification products from damaged templates such as archaeological DNA are made up of a high proportion of chimeric molecules.” EX1019 (Paabo), 1. In other words, Paabo provides only mechanistic insight into PCR recombination and does not discuss or suggest PCR recombination between two different samples.

110. Odelberg provides an “alternative explanation” for PCR recombination events involving “a template-switching process in which the polymerase or the nascent strand switches from the original template to a secondary template during DNA synthesis.” EX1020 (Odelberg), 1. Odelberg does not analyze tagged nucleic acids but shows that

recombination events can arise during “a single primer-extension phase” that is “consistent with a template-switching mechanism in which the polymerase or 3’-end of the nascent strand switches from the original template to the complementary nascent strand that is being generated from the other PCR primer.” EX1020 (Odelberg), 4. In other words, Odelberg provides only further mechanistic insight into PCR recombination and does not suggest PCR recombination between two different samples.

111. Dr. Spellman states that “[a]s 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).” EX1003 (Spellman Declaration), ¶ 68. Both Creer and Dai examine chimeric sequences in next-generation sequencing data but neither describes cross-over errors.

112. Creer describes pyrosequencing of the meiofaunal biosphere and the impact of “putative recombinant DNA molecules (chimeras) in metagenetic data sets.” EX1034 (Creer), 1. Specifically, Creer performs one PCR reaction per sample using tagged primers targeting the 18s rDNA gene and pools the tagged nucleic acids for pyrosequencing. Creer observes many putative chimeric OCTUs (operational clustering of taxonomic units) which they deem “disconcerting” and conclude “[i]n many ways, metagenetic ultrasequencing experiments are the ideal ‘breeding ground’ for recombinant DNA molecules.” EX1034 (Creer), 10. Yet, Creer does not propose that any chimeric sequences have formed between template nucleic acids *of different samples*. Rather, the suggestions they provide refer to the individual “environmental PCRs” *of each sample*. EX1034 (Creer), 10 (“It is therefore better to try and reduce the level of DNA recombination *within environmental PCRs* by adhering to the following procedures.”).

113. Dai describes the “NGSQC (Next Generation Sequencing Quality Control) pipeline” which “enables researchers to determine whether sequencing data related to their most interesting biological discoveries are caused by sequencing quality issues.” EX1033 (Dai), 1. Dai notes

that “[a] key issue in paired-end/mate pair data analysis is the false chimeric sequences, whose two ends are not from the original continuous sequence fragments during library preparation.” EX1033 (Dai), 8. However, Dai does not show whether these are cross-over errors *from different samples*. Dai merely speculates these could be “caused by problems in library preparation” or “the result of poor sequencing library preparation.” EX1033 (Dai), 9. In other words, these could simply be chimeric sequences formed from template nucleic acids derived from library preparation *of each sample*.

114. The literature at the time of the invention explained that cross-over error for flow-cell based NGS was not thought to be a problem. For example, Lefrancois (EX2005) describes “a multiplex barcoding system that allows simultaneous sequencing and analysis of multiple samples using Illumina's platform” “to perform efficient ChIP-Seq in yeast.” EX2005 (Lefrancois), 1. Lefrancois found that “[t]his method produces accurate results with higher throughput and reduced cost.” EX2005 (Lefrancois), 1. Lefrancois defines crossover error as “the erroneous assignment of a barcoded read from a particular ChIP sample to the wrong ChIP sample during multiplex ChIP-Seq, generating a signal profile for one barcoded sample with reads from other ChIPs

sequenced simultaneously.” EX2005 (Lefrancois). In evaluating the results, Lefrancois noted that “analysis of different ChIP samples with distinct barcodes indicates that *the signal from one barcoded library does not crossover to those of other barcoded libraries.*” EX2005 (Lefrancois). Though Lefrancois contemplates crossover error, it determined that *crossover error did not occur* using Illumina’s platform.

115. Additionally, refuting Dr. Spellman’s opinions that a POSA would know that “[p]roblems with single indexes could also arise during PCR amplification when chimeric molecules of the template nucleic acids were formed,” as discussed above in Section X.A, Parameswaran reported the unambiguous assignment of 99.8% of the observed sequences with the sequencing and identification of a *single error-free barcode* at either end of the sequence of interest. Section X.A; EX1004 (Parameswaran), 1. Based on Parameswaran’s disclosure, a POSA would understand that using only single indexes would not create problems with assignment.
116. Aside from a general lack of awareness of cross-over error in NGS machines, a POSA would additionally understand that the mismatched error-free barcode motifs observed in Parameswaran were not a result

of cross-over because the majority of error-free barcode motifs observed were resolved with the addition of a cloning linker to the search motif. Parameswaran included the cloning linker because six of the forward barcodes are exact reverse complements of six reverse barcodes. EX1004 (Parameswaran), 7. A POSA would understand that forward and reverse barcodes that are “exact reverse complements” of one another have an increased risk of mismatched barcodes because a given forward barcode could be incorrectly read as a reverse barcode associated with another sample. Without including a cloning linker in its search motif, Parameswaran observed 2349 sequences with error-free mismatched barcode motifs in Run I, and 5061 sequences with error-free mismatched barcode motifs in Run II. EX1004 (Parameswaran), 7. By including a 3-nucleotide linker in the search motif of its analysis, Parameswaran reported that it only observed 3 sequences with error-free mismatched barcode motifs in Run I, and 11 sequences with error-free mismatched barcode motifs in Run II. EX1004 (Parameswaran), 7. A POSA would understand that cross-over error from different samples would not be reduced with the inclusion of a cloning linker in the data analysis step used to assign sequences to their given sample. Instead, a POSA would understand that cross-over

error from two different samples would remain the same regardless of the search motif data analysis step used to assign sequences to their given sample.

117. Dr. Spellman fails to appreciate that Parameswaran itself did not suggest or imply that cross-over error was occurring. EX1004 (Parameswaran), 7-8. As an initial matter, Parameswaran does not mention “chimeric” or “chimeric molecules” once. EX1004 (Parameswaran). Instead, Dr. Spellman argues that a POSA would have known that mismatched error-free barcode motifs in Parameswaran included chimeric sequences resulting from template cross-over error. EX1003 (Spellman Declaration), ¶¶ 67–69, 147-148, 160-161. But Dr. Spellman only provides support that a POSA would have known chimeric sequences could form during PCR amplification of *individual samples* when homologous sequences are amplified, thus creating chimeric sequences within a single sample. EX1003 (Spellman Declaration), ¶¶ 67–69 (citing Lahr, Meyerhans, Paabo, Creer, and Dai). Dr. Spellman does not show that a POSA would understand that the PCR recombination of template nucleic acids from *different samples* would produce chimeric sequences that could create cross-over error. EX1003 (Spellman Declaration), ¶¶ 67–69.

118. To meet the limitations of Claim 1, Dr. Spellman asserts that “[a] POSA would have understood that the sequences with mismatched barcode motifs in Parameswaran included those resulting from *template cross-over error* introduced during the sequencing workflow (sequences of chimeric molecules *made from template nucleic acids derived from different samples*).” EX1003 (Spellman Declaration), ¶ 130. More specifically, Dr. Spellman asserts that “Parameswaran acknowledges that sequences with *mismatched or misassigned barcode motifs*, the same type of sequences associated with chimerism, can result from various errors including cross-contamination of template DNAs during PCR.” EX1003 (Spellman Declaration), ¶ 130. But Dr. Spellman defines “template crossover error” to mean “sequences of chimeric molecules made from template nucleic acids *derived from different samples*.” EX1003 (Spellman Declaration), ¶ 122. Dr. Spellman fails to show that a POSA would understand the mismatched or misassigned barcode motifs to be chimeric molecules resulting from cross-over error between template nucleic acids *derived from different samples*.

119. To be sure, Parameswaran identifies three potential sources of error that lead to sequences with mismatched barcode motifs – none of which suggest or imply that sequences with mismatched barcode motifs are

indicative of cross-over error. Section X.A.; EX1004 (Parameswaran), 7-8.

120. **First**, Parameswaran suggests that the error-free mismatched barcode motifs could have been caused by errors in barcode synthesis or sequencing. EX1004 (Parameswaran), 7. A POSA would understand that errors in barcode synthesis and sequencing would result in errors within the barcode sequence itself, not chimeric sequences resulting from cross-over events during post-pooling amplification. EX1004 (Parameswaran), 7; EX1003 (Spellman Declaration), ¶ 130. A POSA would understand that errors in barcode synthesis or sequencing could lead to mismatched barcodes if a barcode was incorrectly synthesized or sequenced to match another the sequence of another barcode belong to a different sample. APOSA would not expect errors in barcode synthesis or sequencing to lead to chimeric sequences within template nucleic acids, let alone chimeric sequences between template nucleic acids of *different samples* as required by Dr. Spellman's own definition of cross-over error.

121. A POSA would understand that errors in barcode synthesis or sequencing resulting in mismatched barcode motifs are more prevalent when one barcode has a similar sequence as another barcode. For

example, when a barcode differs from another barcode by a single nucleotide, a single sequencing or synthesis error could lead to sequences with error-free mismatched barcodes instead of an error-containing barcode, or what Parameswaran refers to as an “imperfect” barcode. EX1004 (Parameswaran), 7. Parameswaran therefore designed its barcodes to “differ[] in at least 4nt” and required error-free barcodes for its analysis. EX1004 (Parameswaran), 4.

122. ***Second***, Parameswaran points to cross-contamination in the second round of PCR and contamination in PCR primers as creating sequences with mismatched error-free barcodes. EX1004 (Parameswaran), 7. Neither cross-contamination in the second round of PCR nor contamination in PCR primers would necessarily result in chimeric sequences from different samples, and these would not result in chimeric sequences created during cluster amplification. Cross-contamination in the second round of PCR of Parameswaran involves further amplification of already-barcoded PCR products, using the same set of barcoded primers used in the first round of PCR. EX1004 (Parameswaran), 3, 7. Therefore, if a tagged template DNA spills into a different well, it could produce a sequence with mismatched barcodes simply because it is PCR amplified by an abundance of barcoded

primers intended to tag a different sample. This would not produce a chimeric sequence made from template nucleic acids derived from different samples. In a similar way, contamination of PCR primer stocks would not produce chimeric sequences made from template nucleic acids derived from *different samples*. Instead, both cross-contamination in the second round of PCR and contamination in PCR primer stocks could produce mismatched barcodes throughout the template nucleic acid sequence data from one sample.

123. Additionally, both cross-contamination in the second round of PCR and contamination of PCR primers and are events that occur prior to pooling or sequencing, so Parameswaran further acknowledges that “any possible source for sample contamination upstream (or during) library construction needs to be carefully monitored.” EX1004 (Parameswaran), 7. Therefore, a POSA would understand from Parameswaran that mismatched error-free barcodes would likely have arisen from contamination events occurring upstream of (or during) the library construction of individual samples and not during the pooling, amplification for cluster generation, or sequencing steps.
124. Dr. Spellman also states that Gloor demonstrated that “identification and removal of chimeric sequences was desirable when performing

Illumina sequencing,” (EX1003 (Spellman Declaration), ¶ 161), but as mentioned above, Gloor does not tie any identification of chimeric sequences to cross-over error, nor mention cross-over error at all. Section X.B. Indeed, POSA would understand that Gloor could not identify the prevalence of cross-over error with its combinatorial tagging method. Section X.B. Gloor also found so few chimeric sequences, using a chimerism test separate from its CDI method, that it used its dataset “without regard to chimeric sequences because putative chimeric sequences composed a miniscule fraction of reads.” EX1005 (Gloor), 8, 22. Therefore, Gloor did not acknowledge cross-over error at all or identify error from chimeric molecules as a problem, and a POSA would not have understood Gloor to suggest spending any time or effort identifying or removing chimeric sequences, as Dr. Spellman suggests.

125. Beyond Parameswaran and Gloor, based on the state of the art, a POSA the time of the invention of the '852 patent would have understood that the false assignment rate of NGS sequencers could be sufficiently addressed with a single error-free index. For example, Meyer 2007 uses a barcoding method dubbed “parallel tagged sequencing (PTS), a simple, inexpensive and flexible barcoding technique that can be used

for parallel sequencing any number and type of double-stranded nucleic acid samples.” EX2010 (Meyer 2007), 1. Meyer 2007 appends dual barcodes but only requires a single barcode for read assignment, like a single indexing barcoding method. For instance, they remove the second barcode if it is found in the read. EX2010 (Meyer 2007), 3 (“The 3’ ends of the reads were then scanned for the reverse-complement of the barcode sequence and, if found, the ‘clip_adapter_right’ entry was set to also remove this part.”). Meyer 2007 highlighted the ability to “currently provide sequences for more than 50 mtDNA genomes at ~20-fold coverage,” which is useful because “aiming for 20-fold average coverage is sufficient to obtain complete consensus sequences for all or the vast majority of the samples.” EX2010 (Meyer 2007), 3. Meyer 2007 explains that “errors can be easily identified and eliminated when closely related sequences are available for comparison” with the single-index PTS method described. EX2010 (Meyer 2007), 3.

126. Wang similarly describes a “robust method of generating strand-specific Illumina-compatible libraries for RNAseq,” where dual indexes are added but only single indexes were analyzed using single-end sequencing. EX2011 (Wang), 2. Wang not only used in effect single indexes without mentioning the possibility of cross-over error,

but allowed reads into its dataset that included one sequencing error in the first five base pairs. EX2011 (Wang), 6. Wang reasoned that “by including reads with one mismatch to the index, it was possible to reclaim an additional 5% of the total mappable reads (4.8 million reads).” And Kenny describes “an advanced method that combines DNA samples using indexes or barcodes” “to enrich the [DNA] samples in parallel.” EX2012 (Kenny), 1. Kenny’s single-indexing barcoding method, with “individual DNA samples, each incorporating a unique 6-bp index” resulted in “99% of sequence SNP calls were concordant with known genotypes.” EX2012 (Kenny), 1.

127. Additionally, even Binladen, a reference Dr. Spellman relies on, uses single-index analysis without consideration of or concern about cross-over error. Binladen describes using “5’-nucleotide tagged primers to generate homologous DNA amplification products.” EX1029 (Binladen), 1. Binladen attached the same barcode to either end of its nucleic acid template so that the sequencer would be sure to read one tag despite limitations in read length. EX1029 (Binladen), 3 (“our analyses were limited to simply discriminating using the primer at the sequence end of the product.”). Binladen even recognized that full sequences with tag sequences on both ends could not be observed. EX1029

(Binladen), 3 (“However, under the current status of the sequencing technology, GS20 sequencing reads are limited to approximately 120 bases, thus in *this experiment the full sequence (133-141 bp including primer, species dependent) was not returned.*”). Binladen highlighted that its indexing method “enables accurate sequencing and assignment of homologous DNA sequences from multiple sources in single high-throughput GS20 run” based on the presence of a single tag sequence, without any acknowledgement of potential cross-over error. EX1029 (Binladen), 1. Therefore, even with the existence of dual-index methods, a POSA would not have known that crossover error was a problem in NGS.

128. A POSA also would not have been motivated to identify and discard cross-over error because Illumina itself provided no indication that it was aware of cross-over error, or that it was an issue on Illumina sequencers. Indeed, Illumina’s own documents around the relevant time advertised the use of single indexes with its sequencing systems. For example, Illumina touted the ability to multiplex sequence using a unique single index in its 2008 publication titled “Multiplexed Sequencing with Illumina Genome Analyzer System.” EX2007 (Multiplexed sequencing with Illumina Genome Analyzer System), 1

(“In the multiplexed sequencing method, DNA libraries are ‘tagged’ with *a unique identifier, or index, during sample preparation.*”), 2 (“To identify samples after pooling, each sample is uniquely tagged with a sequence index during the sample preparation protocol.... The Multiplexing Sample Preparation Oligonucleotide Kit provides *12 index oligos* for pooling *up to 12 samples per lane.*”). Illumina also proposed a single indexing method in a sequencing application note titled “High-Speed, Multiplexed 16S Microbial Sequencing on the MiSeq® System” around 2011, 2012 and 2014. EX2006 (High-Speed, Multiplexed 16S Microbial Sequencing on the MiSeq System), 1 (bearing 2011-2012, 2014 copyright date). A POSA would understand that Illumina’s application note provides a single indexing method to multiplex 16S microbial communities using Illumina sequencers. EX2006 (High-Speed, Multiplexed 16S Microbial Sequencing on the MiSeq System), 1 (Figure 1 showing a single index), 1-2 (“The V4 region from 24 samples was amplified using primer encoding 24 different barcodes and combined into a single library on the MiSeq system.”). A POSA would understand that Illumina’s own documents suggested single indexing methods for Illumina Genome Analyzers and

sequencing 16S microbial communities around the time of the '852 patent.

129. Illumina did not identify cross-over error as an issue on its own Illumina sequencers in a publication, or otherwise, until the white paper titled “Effects of Index Misassignment on Multiplexing and Downstream Analysis” around 2018. EX2003 (UDI White Paper), 1 (bearing 2018 copyright date). In its white paper, Illumina explains “[t]he unique dual indexes eliminate hopped reads from downstream analysis, as unexpected combinations are assigned as undetermined and removed from the data.” EX2003 (UDI White Paper), 3. A POSA would understand that Illumina did not identify cross-over error and therefore did not suggest finding any method to identify and reduce such error, nor suggest using unique dual indexes, until its white paper.
130. Consistent with the lack of disclosure in Parameswaran, Gloor, and the art at the time of the invention, Dr. Spellman has not shown that a POSA would have been aware of cross-over error due to chimeric sequences generated in cluster amplification at all at the time of the '852 Patent, let alone be motivated to identify such error and discard chimeric sequences to reduce it.

**iii. Dr. Spellman Does Not Show That
Parameswaran's Speculation Regarding Other
Sequencers Would Have Motivated The
Proposed Combination**

131. Dr. Spellman notes Parameswaran mentioned other sequencing platforms that were available, but Parameswaran provides no details regarding such systems or how its methods would be implemented on them. EX1003 (Spellman Declaration), ¶¶ 162-163; EX1004 (Parameswaran), 2, 8. As explained above in Section XI.A.1., a POSA reading Parameswaran would understand it to be suggesting that the use of its barcodes may enable multiplexing on other sequencing platforms, but would also understand, by the time of the invention, that the field had moved beyond Parameswaran's methods and already enabled multiplexing in Illumina sequencers. As also explained above, a POSA would not understand Parameswaran to have any utility over already robust methods for multiplexing on Illumina sequencers, such as Gloor. Dr. Spellman also states that Maughan 2010 (EX1038) would have reinforced motivation to combine. EX1003 (Spellman Declaration), ¶ 163. But Maughan 2010, like Parameswaran, used a dual-indexing method that assigned sequences to the correct samples based on the presence of a single barcode sequence. EX1038 (Maughan 2010), 5-6. Therefore, neither Parameswaran nor Maughan 2010 would motivate a

POSA to simply try out unique dual indexes as an alternative to Gloor's sequencing method, which already required dual indexes.

B. A POSA Would Not Have Reasonably Expected Success In Combining Parameswaran With Gloor

132. Dr. Spellman states that “a POSA in December 2010 would have had a reasonable expectation of success in combining *Parameswaran* and *Gloor* to perform the claimed methods.” EX1003 (Spellman Declaration), ¶ 165. However, the claims of the '852 Patent require “a method for reducing template cross-over error” and actions taken “to reduce cross-over error,” and Dr. Spellman does not show why or how a POSA would reasonably expect to reduce cross-over error using Parameswaran's UDIs with Gloor's Illumina sequencing method. A POSA would not expect unique dual indexes to reduce crossover error because a POSA would not realize that crossover error was occurring.
133. As discussed above, a POSA in 2010 would not have understood the prevalence of crossover error arising after tagged amplicons were pooled for sequencing during amplification to form clusters, either on the surface of beads in the case of Roche 454 sequencing or on the surface of the flow cell for Illumina sequencing. *See* Section XI.A.2, above.

134. As also discussed above, Dr. Spellman did not identify any reference, including Parameswaran and Gloor, that shows crossover error arising from chimeric molecules formed during amplification after sample pooling was a known issue for POSAs at the time of the invention. For example, the exhibits provided by Dr. Spellman demonstrate that recombination could arise between template nucleic acids of a single PCR reaction. Yet, even under Dr. Spellman's definition of cross-over error, those references do not disclose such error as none of them show that PCR recombination can occur between template nucleic acids of *different samples*. See Section XI.A.2, above; EX1003 (Spellman Declaration), ¶¶ 67–69. Further, because PCR amplification using tagged primers would be generally performed as a closed tube reaction of *a single sample per PCR reaction*, to assign sequence tags to each sample, a POSA would not expect the recombination of nucleic acids between samples. See Section XI.A.2, above.

135. Dr. Spellman states that a POSA would have understood that they could have combined the unique-dual indexing method of Parameswaran with an Illumina sequencer without any “novel or nonobvious chemistry,” and expected success. EX1003 (Spellman Declaration), ¶¶ 166–67. However, Dr. Spellman does not point to anything to say that a POSA

would have reasonably expected Parameswaran's method to reduce cross-over error according to his definition, either on the pyrosequencing platform on which it was performed or other sequencing platforms. And Dr. Spellman does not point to anything in Parameswaran to show why a POSA would have considered using its method to identify cross-over error from chimeric molecules formed during amplification after sample pooling. In fact, Parameswaran does not discuss the existence of cross-over at all, let alone that its methods could be used to reduce sample identification issues arising from these errors. As I further explain above in Section X.B, especially considering that a POSA would have been unaware of cross-over error as described in the '852 Patent, a POSA would not read Parameswaran to suggest that its method could be used to reduce the error due to formation of chimeric molecules during amplification on the surface of flow cells. A POSA would therefore not reasonably expect that Parameswaran's method would reduce such a problem.

136. To the contrary, Parameswaran taught that it could "unambiguously assign 99.8% of obtained barcoded sequences by bioinformatically probing for an error-free barcode at *either end of the sequence of interest*" and that its method had a "remarkably low misassignment

rate” by using its method on a Roche 454 pyrosequencer. EX1004, 2, 8. Because the method worked well, nothing in Parameswaran suggests doing anything to further reduce other unidentified errors, such as cross-over errors arising during amplification, which were not even recognized as an issue at the time of the invention. EX2004 (Kircher 2011); EX2003 (Illumina UDI White Paper). In fact, Parameswaran explicitly attributed any errors giving rise to “perfect mismatched barcodes” as attributable to contamination in PCR amplification of each sample, barcode synthesis, and barcode sequencing, not cross-over error between molecules during cluster amplification after sample pooling. EX1004, 7.

137. In terms of its application with other sequencers, Parameswaran only says that its barcodes could “be compatible with other platforms,” not that they would reduce any error, let alone cross-over error if used. EX1004, 8. To the contrary, Parameswaran explains that if using its barcodes on other systems, the “structure of the barcodes might need to be revised to reflect sequencing error profiles of the various systems.” EX1004, 8. Therefore, although Parameswaran concluded that its method worked well with pyrosequencers, it did not provide any reason for a POSA to expect that it could work well with other sequencers, like

an Illumina flow-cell. Based on the disclosure in Parameswaran, a POSA would expect that it would need to make changes to Parameswaran's barcodes to address Illumina specific errors. Dr. Spellman does not suggest any such modifications and does not provide any reason that a POSA would know to address cross-over error when a POSA at the time would not have appreciated the existence of cross-over error in Illumina sequencers.

138. Additionally, Gloor shows that its combinatorial tagging method on Illumina permits more accurate microbiome profiling at a lower cost than pyrosequencing. For example, Gloor describes that "...pyrosequencing is constrained by cost limitations and a relatively high per-read error rate," that "[t]he 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," and that "the *higher fidelity Illumina sequencing* may have resulted in fewer taxa because of a *lower error rate* contributing to fewer 'rare microbiome' taxa." EX1005, 23. Parameswaran suggested that the benefits Gloor touts would be possible, concluding that because of the low misassignment rate, combinatorial tagging could permit exponentially greater sample

throughput with a given set of tags, as Gloor later used. EX1004, 8 (“The low misassignment rate also indicates the potential utility of combinations of forward and reverse barcodes to generate up to 48^2 (=2304) combination barcodes for this set of sequences.”). In other words, Gloor provides an improvement for microbiome profiling over Parameswaran, using a method suggested by Parameswaran, and was successful. Dr. Spellman does not identify anything in Parameswaran, Gloor, or the art, that would provide a POSA with an expectation that using Parameswaran’s UDIs would provide any benefit over Gloor’s CDI method. As I explain above in Section XI.A.2.i, a POSA would not think there was any utility to using UDI over CDI on Illumina’s sequencers, and would instead think that implementing Parameswaran’s pyrosequencing-tailored UDI method would simply lead to more expense and less efficiency.

C. Dr. Spellman Fails To Show That The Combination Discloses Or Renders Obvious “Discarding Sequence Reads Obtained In Said Sequencing Step That Contain Said Combination Of Identifier Sequences That Are Different Than Any Of Said Distinct Pairs, Thereby To Reduce Cross-Over Error.”

139. Dr. Spellman asserts that Parameswaran discloses the discarding limitation of the ’852 patent. EX1003 (Spellman Declaration), ¶ 152. Specifically, Dr. Spellman says that, because Parameswaran identified

sequences with error-free mismatched barcode motifs, a POSA would understand that Parameswaran “disclosed a method in which [] chimeric sequences could be discarded, and thus, reduce cross-over error.” EX1003 (Spellman Declaration), ¶ 153. Dr. Spellman misreads Parameswaran and the ’852 patent.

140. The discarding limitation of the ’852 patent requires actually “*discarding* sequence reads...that contain said *combination of identifier sequences that are different than any of said distinct pairs.*” EX1001, 21:16-19. However, Dr. Spellman only points to Parameswaran’s *identification* of sequences with error-free mismatched barcode motifs as support to say that Parameswaran discloses discarding sequence reads with a mismatched “combination of identifier sequences.” EX1003 (Spellman Declaration), ¶¶ 153–54. But, as Dr. Spellman recognizes, Parameswaran does not discard these sequences in its analysis. EX1004 (Parameswaran), 7; EX1003 (Spellman Declaration), ¶ 153. Indeed, Parameswaran still kept sequences with incorrect barcode pairs as long as they contained at least one sequence with a match to one of the two expected barcode motifs. EX1004 (Parameswaran), 8. Put differently, Parameswaran only stated excluding sequences that had *no* perfect forward *or* reverse barcodes.

EX1004 (Parameswaran), 8. Parameswaran reasoned that its method would not cause any issues with placing the sequences into the correct bin with reasonable accuracy given the very low false discovery rate of sequences with perfect and unexpected barcode motifs (i.e., less than 0.005%) when 13nt barcode motifs were used and would still allow them to maximize the number of sequence reads. EX1004 (Parameswaran), 8. Because Parameswaran decided to keep these sequence reads containing at least one perfect barcode motif, Dr. Spellman is incorrect that it discloses discarding sequence reads with a mismatched “combination” of barcode motifs.

141. Dr. Spellman also states that a POSA would have understood that error-free mismatched barcode motifs “*could* be excluded to increase accuracy” because “Parameswaran states that ‘any sequence with both 5’ and 3’ barcode motifs will be placed with reasonable accuracy into the correct bin.” EX1003 (Spellman Declaration), ¶ 154 (citing EX1004 (Parameswaran), 7-8). But Dr. Spellman again confuses Parameswaran. Parameswaran’s full statement merely explains that “[t]he low false-discovery rate 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.

It does not suggest to a POSA that they could or should discard sequences with perfect and unexpected barcode motifs.

142. Because of the low misassignment rate, Parameswaran also suggests using combinatorial tagging, which could be used to sequence more samples with fewer barcodes but would increase the possibility of undetectable, incorrect barcode pairs. EX1004 (Parameswaran), 8. This suggestion would further direct a POSA away from thinking they could or should discard sequences with an incorrect “combination” of barcode motifs, because combinatorial indexing would not allow them to even determine if a seemingly correct pair were actually traceable to a sample.

143. In fact, Parameswaran described the success of its methods when performed as is, noting they obtained “[a] remarkably low misassignment rate (0.00042% and 0.0044% for two independent experiments)” by using 13nt barcode motifs and discarding only those without any perfect forward or reverse barcodes while retaining reads with at least one perfect forward or reverse barcode. EX1004 (Parameswaran), 8. From this and the low-false discovery rate at which mismatched error-free barcode motifs were actually observed, a POSA would not see a need to identify and discard sequences containing a

mismatched pair of barcode motifs. A POSA would further understand that keeping sequences with either error-free barcode motif would allow them to maximize the number of sequence reads included in their analysis.

144. Thus, a POSA would not have understood Parameswaran as disclosing a method to discard mismatched error-free barcode motifs, nor would a POSA find it obvious that they “could” or should exclude such sequences. EX1003 (Spellman Declaration), ¶ 154. Instead, a POSA would think that, because of Parameswaran’s low false-discovery rate of mismatched perfect barcodes, such sequences could be included in the analysis so long as at least one barcode were error-free. In effect, Parameswaran discloses nothing more than the methods used before and identified by the ’852 Patent (i.e., use a single identifier and only discard if there is not any identifier associated with the molecule for tracing it back to a sample), which, as I explain above, are not sufficient for detecting chimeric molecules formed during amplification. Section VIII (above); EX1001, Fig. 1, Fig. 3, 4:57-63, 6:17-20.

145. Dr. Spellman also states that Gloor discloses the discarding step because Gloor explained that “many chimeric sequences *may have been removed* because of the strict requirement for proper sequence tag

and primer sequences on the left and right ends.” EX1003 (Spellman Declaration), ¶ 155 (citing EX1005, 8, 22-23). But neither Dr. Spellman nor Gloor identify any excluded or discarded sequence that is a chimeric molecule containing a “combination of identifier sequences that are different than any of said distinct pairs.” *See* Section X.B (above).

146. For example, Dr. Spellman does not explain how the “chimeric sequences” allegedly excluded in Gloor are formed during amplification after sample pooling, like the sequences that give rise to cross-over error in the ’852 Patent. EX1003 (Spellman Declaration), ¶ 136 (“By removing (i.e., discarding) the chimeric sequences, Gloor teaches a method of reducing cross-over error introduced during sequencing workflow.”). Dr. Spellman recognizes that, in Gloor, “putative chimeric sequences” reflect “rare, incorrect sequences made of two abundant, correct sequences” EX1003 (Spellman Declaration), ¶ 98 and “true “chimerism”” constitutes sequences “made of two different bacterial reference sequences” EX1003 (Spellman Declaration), ¶ 98. But Gloor does not specify whether these incorrect chimeric sequences arose during amplification after sample pooling, nor does Gloor say whether these chimeric sequences contain

mismatched tags. EX1005 (Gloor), 8. Indeed, a POSA would understand that incorrect chimeric sequences are not necessarily derived from template nucleic acids of two or more samples and can contain two correct tags. For example, Dr. Spellman acknowledges that incorrect chimeric sequences can arise from PCR amplification of a single sample without the potential read misassignment. EX1003 (Spellman Declaration), ¶ 67.

147. In fact, Gloor explains that there would have been “very few chimeric sequences” left after using their method because they used only a small amount of DNA and few amplification cycles, reducing the risk of chimeric molecule formation. EX1005 (Gloor), 22. And Gloor explains that “many chimeric sequences *may have been removed*” by overlapping (disconnected from any sequence tag analysis) because prior to analysis of the sequence tags, only sequence reads with a perfectly overlapping region were further included for later analysis. EX1005 (Gloor), Fig. 2, 22. This means that only those molecules that met the “requirement for long overlapping segments of a defined length” (i.e., those that did not display any mismatch in the paired-end reads) were kept for later analysis. EX1005 (Gloor), Fig. 2, 22. As a result of this requirement, chimeric molecules would have been

discarded without any analysis of whether they contained a correct pair of sequence tags. That makes sense, as Dr. Spellman recognizes, because Gloor uses “non-unique dual indexing” (i.e., combinatorial indexing), and thus it cannot “*discard*” sequence reads...that contain said *combination of identifier sequences that are different than any of said distinct pairs*,” because the combination of identifier sequences associated with a chimeric molecule formed from a cross-over error could be the same as another valid pair belonging to a different sample. EX1003 (Spellman Declaration), ¶ 155. Gloor’s non-unique dual indexing method would result, then, in assigning sequence reads from a chimeric molecule to a sample instead of discarding them. From this, a POSA would understand that Gloor does not disclose how those sequences would have been identified as having an incorrect “combination of identifier sequences” different from any distinct pairs. EX1005 (Gloor), 22.

148. Because neither Parameswaran nor Gloor disclose or render obvious “discarding sequence reads ... that contain said combination of identifier sequences that are different than any of said distinct pairs,” this limitation is not met in the combination.

**XII. GROUND 2: THE CHALLENGED CLAIMS ARE NOT OBVIOUS
OVER PARAMESWARAN AND BENTLEY**

**A. A POSA Would Not Be Motivated To Combine Parameswaran
And Bentley.**

149. Dr. Spellman states that a POSA would combine Parameswaran and Bentley for many of the same reasons that I address above for Parameswaran and Gloor. EX1003 (Spellman Declaration), ¶ 207. For the same reasons I discuss above regarding why a POSA would not have combined Parameswaran with Gloor's Illumina sequencers, a POSA similarly would not combine Parameswaran with Bentley's Illumina sequencers. *See* Section XI.A (above). For example, Dr. Spellman states that a POSA would have understood Bentley's Illumina sequencers as being more efficient and costing less, that Parameswaran's method would improve the "already superior accuracy" of Illumina Sequencers, and that Parameswaran suggests using its method on other sequencing platforms. EX1003 (Spellman Declaration), ¶ 207. But Bentley does not have any discussion of multiplexing. Instead, Bentley performed whole-genome sequencing for a sample using Illumina sequencers. EX1006 (Bentley), 1.
150. Additionally, Bentley describes the mechanisms in Illumina sequencers that use bridging amplification to form DNA clusters, but Bentley does

not suggest combining multiple samples together. EX1006 (Bentley),

1. In fact, Bentley explains that the novelty of its approach was more efficiently determining “the genome sequence of *a male Yoruba* from Ibadan, Nigeria,” meaning sample identification was not an issue Bentley addressed—all sequences were from the same individual.

EX1006 (Bentley), 1. Bentley also does not speculate about cross-over errors arising during amplification on flow-cells, nor does it even mention the possibility of chimeric molecules forming during its sequencing workflow. This is not surprising, as the industry did not recognize the prevalence of chimeric sequences formed in cluster amplification, and whole-genome sequencing using Illumina flow-cells was in its infancy. EX2004 (Kircher 2011); EX2003 (Illumina UDI White Paper). Dr. Spellman does not point to anything to explain why a POSA would be motivated to use Parameswaran’s sample identifiers in Bentley when sample identification was not even at issue, and a POSA would not have been motivated to combine the two references.

B. A POSA Would Not Have Reasonably Expected Success In Combining Parameswaran and Bentley.

151. Dr. Spellman states that a POSA would have expected success in combining Parameswaran’s unique dual-indexing method with Bentley’s whole-genome sequencing for many of the same reasons that

I address above for Parameswaran and Gloor. EX1003 (Spellman Declaration), ¶¶ 208-212. For the same reasons I discuss above regarding why a POSA would not have reasonably expected success in combining Parameswaran's unique dual indexing with Gloor's Illumina sequencers, a POSA similarly would not have expected success in the combination of Parameswaran and Bentley. *See* Section XI.B (above).

C. Dr. Spellman Fails To Show That The Combination Discloses Or Renders Obvious “Discarding Sequence Reads Obtained In Said Sequencing Step That Contain Said Combination Of Identifier Sequences That Are Different Than Any Of Said Distinct Pairs, Thereby To Reduce Cross-Over Error.”

152. As I describe above in Section XI.C, Parameswaran does not disclose “discarding sequence reads” containing a “combination of identifier sequences that are different than any of said distinct pairs.” Bentley does not perform multiplexing of multiple samples, does not discuss chimeric sequences, does not describe cross-over errors, and does not disclose identifying nucleic acids such that sequence reads from chimeric molecules would be identified and discarded due to an incorrect combination of identifier sequences. EX1006 (Bentley), 1. Instead, Bentley only discloses sequencing the whole genome of a single individual. Petitioner and Dr. Spellman thus fail to point to any

part of Parameswaran or Bentley to disclose this limitation and thus do not make this step obvious.

XIII. APPENDIX A: THE CHALLENGED CLAIMS OF THE '852 PATENT

1. A method for reducing template cross-over error introduced during sequencing workflow, the method comprising the steps of

incorporating at least two members of a plurality of identifier sequences into template nucleic acids obtained from at least two different samples, wherein said two members constitute a distinct pair associated with said template;

combining said nucleic acid templates into a single sample;

amplifying said nucleic acid templates on a surface of a flow cell thereby to form clusters, at least one of which comprises a chimeric sequence comprising a combination of identifier sequences that are different than any of said distinct pairs;

sequencing amplicons obtained in said amplifying step; and

discarding sequence reads obtained in said sequencing step that contain said combination of identifier sequences that are different than any of said distinct pairs, thereby to reduce cross-over error.

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

3. The method of claim 1, wherein the identifiers are first and second barcode sequences.

4. The method of claim 1, wherein a first identifier sequence is incorporated into a 5' portion of the template, and a second identifier sequence is incorporated into a 3' portion of the template.

5. The method of claim 4, wherein after incorporation of the identifier sequences, the template has the following configuration from 5' to 3': a first known sequence; a first barcode sequence; a second known sequence; a target polynucleotide sequence; a third known sequence; a second barcode sequence; and a fourth known sequence.

6. The method of claim 1, wherein a first identifier sequence is incorporated into a 5' end of the template, and a second identifier sequence is incorporated into a 3' end of the template.

Case 2024-00964

U.S. Patent No. 11,041,852

7. The method of claim 1, wherein prior to said sequencing step, the template is attached to a substrate.

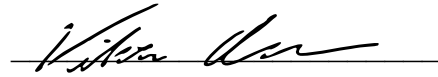
8. The method of claim 7, wherein the template is directly attached to the substrate.

Case IPR2024-00964
U.S. Patent No. 11,041,852

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

April 17, 2025

Date



Dr. Viktor Adalsteinsson, Ph.D.

VIKTOR ADALSTEINSSON, Ph.D.

Greater Boston Area

Phone: (610) 742-7291 Email: viktor.adalsteinsson@gmail.com

EXECUTIVE SUMMARY

Principal Investigator, Innovator, and Entrepreneur, with deep expertise in molecular biology and genomics, developing the next frontier of cancer diagnostics for cancer patients.

RELEVANT EXPERIENCE

Co-Founder and Advisor , Amplifyer Bio	Apr '23 - present
Developing the first "priming agent" injections that are given prior to a blood draw to enhance liquid biopsy testing (see Martin-Alonso et al <i>Science</i> 2024)	
Director , Gerstner Center for Cancer Diagnostics, Broad Institute of MIT and Harvard	Sept '22 - present
Lead a multidisciplinary team of scientists, engineers, computational biologists, and medical oncologists, developing the next frontier of cancer diagnostics	
Associate Director , Gerstner Center for Cancer Diagnostics, Broad Institute	Jan '19 - Sept '22
- Invented multiple technologies for mutation enrichment, DNA repair, linked duplex sequencing, etc., and planned licensing and sponsored research deals - Mentored 20+ team members at varied career stages since establishing the lab	
Group Leader , Blood Biopsy Team, Broad Institute of MIT and Harvard	Jul '16 - Jan '19
- Transferred technologies we developed for scale-up at our genomics platform - Oversaw testing of >10,000 cfDNA samples from patients with >12 cancer types - Analyzed liquid biopsies from a wide variety of translational studies (see papers)	
Research Scientist , Blood Biopsy Team, Broad Institute of MIT and Harvard	Jul '15 - Jul '16
Established the Blood Biopsy Team, a highly collaborative team spanning the Broad Institute, MIT, Harvard, and the Harvard-affiliated hospitals	
Postdoctoral Associate , J. Christopher Love Lab at Koch Institute of MIT	Apr '15 - Jun '15
PhD Candidate , J. Christopher Love Lab at Koch Institute of MIT	Jan '11 - Apr '15
- Developed single-cell assays to measure protein secretion, growth kinetics for drug screening, cell interactions, e.g. B cells secreting tumor-directed Ab's, etc. - Established high-throughput methods for isolation and sequencing of single cells - Demonstrated that scalable whole-exome and -genome sequencing of CTCs is feasible and reflects a matched tumor biopsy	

EDUCATION

Massachusetts Institute of Technology	Aug '10 - Apr '15
PhD, Chemical Engineering	
Dissertation: Genome sequencing and phenotypic analysis of single cells in cancer	
Schreyer Honors College at Penn State University	Aug '06 - May '10
BS, Chemical Engineering	
Dissertation: Ultrasensitive biomolecular detection enabled by nanomaterials	

SELECTED HONORS

Clinical Omics "Pioneers Under 40"	May '21
MIT Technology Review "Innovators under 35"	Aug '17

SELECTED PUBLICATIONS

- C. Martin-Alonso ... V.A. Adalsteinsson, Priming agents transiently reduce the clearance of cell-free DNA to improve liquid biopsies, **Science** 2024.
- T. Blewett ... V.A. Adalsteinsson, MAESTRO-Pool enables highly parallel and specific mutation-enrichment sequencing for minimal residual disease detection in cohort studies. **Clinical Chemistry** 2024.
- J. Bae ... V.A. Adalsteinsson, Single duplex DNA sequencing with CODEC detects mutations with high sensitivity, **Nature Genetics** 2023.
- H. Parsons ... V.A. Adalsteinsson, Circulating tumor DNA association with residual cancer burden after neoadjuvant chemotherapy in triple-negative breast cancer in TBCRC030, **Annals of Oncology** 2023.
- G. Gydush ... V.A. Adalsteinsson, Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth, **Nature Biomedical Engineering** 2022.
- K. Xiong ... V.A. Adalsteinsson, Duplex-Repair enables highly accurate sequencing, despite DNA damage, **Nucleic Acids Research** 2022.
- H.A. Parsons ... V.A. Adalsteinsson, Sensitive detection of minimal residual disease in patients treated for early-stage breast cancer, **Clinical Cancer Research** 2020.
- D.G. Stover ... V.A. Adalsteinsson. Genome-wide copy number analysis of chemotherapy-resistant metastatic triple-negative breast cancer from cell-free DNA. **Journal of Clinical Oncology** 2017.
- V.A. Adalsteinsson ... M. Meyerson, Scalable approach for whole-exome sequencing of cell-free DNA from patients with metastatic cancer, **Nature Communications** 2017.
- J.G. Lohr, V.A. Adalsteinsson ... J.C. Love, Whole exome sequencing of CTCs as a window into metastatic prostate cancer, **Nature Biotechnology** 2014.

For a complete list of publications (>50 manuscripts, >5000 citations), please see <http://pubmed.ncbi.nlm.nih.gov>

SELECTED INVENTIONS

- Minor allele enrichment sequencing through recognition oligonucleotides (US 17/792,638)
- Methods and compositions for detection of tumor DNA (US 18/285,103)
- Compositions and methods for molecular barcoding of DNA molecules prior to mutation enrichment and/or mutation detection (US 11,371,090)
- Methods for genome characterization (US 11,479,878)
- Methods for genome characterization (US 18/463,697)
- Methods for duplex repair (US 18/266,555)
- Methods for duplex sequencing (US 18/266,566)
- Compositions and methods for detecting homologous recombination (WO 2024/124181)
- Methods and compositions for increasing the concentration of tumor DNA (US 18/273,922)
- Size selective tagging of cells in a microfluidic device (US 10,124,336)
- Selective delivery of materials to cells (US 11,806,714)
- Systems, methods, and apparatus for *in vitro* single-cell identification and recovery (US 10,776,608)
- Systems and methods for detection of residual disease (US 18/133,524)

SELECTED CONTRIBUTIONS TO SCIENCE

1. Minimal residual disease (MRD) is the cancer left behind after cancer treatment which could lead to recurrence if left untreated. MRD is difficult to detect with tools such as imaging scans and serum protein biomarkers and being able to detect MRD could inform more precise care for cancer patients. Liquid biopsies have shown promise to detect MRD by assaying tumor mutations in cfDNA but have insufficient sensitivity. Our team has developed ultrasensitive approaches for MRD detection using bespoke assays and mutation enrichment sequencing to track thousands of individualized tumor mutations in cfDNA. We have applied these approaches to over 1000 cfDNA samples from patients treated for early-stage cancer and found that we can detect MRD missed by existing assays.

- G. Gydush, E. Nguyen, J.H. Bae, T. Blewett, J. Rhoades, S.C. Reed, D. Shea, K. Xiong, R. Liu, F. Yu, K.W. Leong, A.D. Choudhury, D.G. Stover, S.M. Tolane, I.E. Krop, J.C. Love, H.A. Parsons, G.M. Makrigiorgos, T.R.

Golub, **V.A. Adalsteinsson**. Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth. *Nature Biomedical Engineering*, 2022. PubMed PMID: 35301450.

- H.A. Parsons, J. Rhoades, S.C. Reed, G. Gydush, P. Ram, P. Exman, K.Xiong, C.C. Lo, T. Li, M. Flehary, G.J. Kirkner, D. Rotem, O. Cohen, F. Yu, M. Fitarelli-Kiehl, K.W. Leong, M.E. Hughes, S.M. Rosenberg, L.C. Collins, K.D. Miller, B. Blumenstiel, L. Trippa, C. Cibulskis, D.S. Neuberg, M. DeFelice, S.S. Freeman, N.J. Lennon, N. Wagle, G. Ha, D.G. Stover, A.D. Choudhury, G. Getz, E.P.Winer, M. Meyerson, N.U. Lin, I. Krop, J.C. Love, G.M. Makrigiorgos, A.H. Partridge, E.L. Mayer, T.R. Golub, **V.A. Adalsteinsson**. Sensitive Detection of Minimal Residual Disease in Patients Treated for Early-Stage Breast Cancer. *Clinical Cancer Research*, 2020. PubMed Central PMCID: PMC7654718

2. Liquid biopsies hold tremendous promise for improving cancer screening, MRD detection, monitoring of tumor evolution, and therapy selection, but lack sensitivity due to limited tumor cfDNA in circulation. Our team has created the first “priming agents” for liquid biopsies which are injections given before a blood draw to enable more cfDNA to be recovered. Whereas most of the field is focused *ex vivo* on sequencing technology, our approaches transiently impair clearance of cfDNA by macrophages and nucleases and improve liquid biopsy cancer detection.

- C. Martin-Alonso, S. Tabrizi*, K. Xiong, T. Blewett, S. Patel, Z. An, S. Sridhar, A. Bekdemir, D. Shea, A. Amini, S. Wang, J. Kirkpatrick, J. Rhoades, T. Golub, J. Love*, S. Bhatia*, **V. Adalsteinsson***. Priming agents transiently reduce the clearance of cell-free DNA to improve liquid biopsies. *Science*, 2024.

3. Next generation sequencing (NGS) has profoundly enabled biomedical research and clinical diagnostic testing. Yet, despite massive improvements in throughput, NGS accuracy remains limited for important applications such as rare mutation detection. Our team has invented multiple methods to improve NGS accuracy such as CODEC, a technique to link both strands of each DNA duplex together so they can be sequenced together, and Duplex-Repair, a technique to prepare DNA for sequencing without substantially resynthesizing DNA duplexes. We have shown that these methods enable >1,500-fold higher accuracy for massively parallel sequencing of single DNA duplexes. We have found that CODEC and Duplex-Repair sequencing of normal blood cells shows an association in mutation load versus age from whole-genome sequencing that is otherwise unapparent in standard NGS. We have also shown that CODEC and Duplex-Repair could improve liquid biopsy sequencing, tumor biopsy sequencing, and detection of mutation signatures such as microsatellite instability (MSI), homologous repair deficiency (HRD), and more.

- J. Bae, R. Liu, E. Nguyen, J. Rhoades, T. Blewett, K. Xiong, D. Shea, G. Gydush, S. Tabrizi, Z. An, S. Patel, G. Makrigiorgos, T. Golub, **V. Adalsteinsson**. CODEC enables ‘single duplex’ sequencing. *Nature Genetics*, 2023.
- K. Xiong, D. Shea, J. Rhoades, T. Blewett, R. Liu, J.H. Bae, E. Nguyen, G. Makrigiorgos, T. Golub, **V. Adalsteinsson**. *Nucleic Acids Research*, 2022. PubMed Central PMCID: PMC8755016.

4. Tumor evolution poses major challenges for cancer care and is difficult to monitor given the inaccessibility of surgical biopsies. Liquid biopsies could overcome this challenge by enabling minimally invasive profiling of patients’ tumors from the cell-free DNA (cfDNA) in blood. Our team completed some of the most comprehensive comparisons of whole-exome sequencing (WES) of cfDNA and matched tumor biopsies, demonstrating high concordance in somatic mutations, copy number alterations, mutation signatures, and even predicted neoantigens. We also established a scalable approach for WES of cfDNA, starting with ultra-low-pass whole-genome sequencing (ULP-WGS) to quantify tumor content in cfDNA. This allowed us to identify the samples with sufficient tumor cfDNA for WES, thereby making it scalable and cost effective. To date, we have applied this approach to over 15,000 blood samples from patients with advanced cancer. Our findings suggest that ULP-WGS and WES could enable routine genomic profiling of patients’ tumors, including when tissue biopsies would otherwise be infeasible.

- D.G. Stover, H.A. Parsons, G. Ha, S. Freeman, W. Barry, H. Guo, A. Choudhury, G. Gydush, S. Reed, J. Rhoades, D. Rotem, M.E. Hughes, D.A. Dillon, A.H. Partridge, N. Wagle, I.E. Krop, G. Getz, T.R. Golub, J.C. Love, E.P. Winer, S.M. Tolane, N.U. Lin, **V.A. Adalsteinsson**. Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. *Journal of Clinical Oncology*, 2018. PubMed Central PMCID: PMC5815405

- **V.A. Adalsteinsson**, G. Ha, S.S. Freeman, A.D. Choudhury, D.G. Stover, H.A. Parsons, G. Gydush, S. Reed, D. Loginov, D. Livitz, D. Rosebrock, I. Leshchiner, J. Kim, C. Stewart, M. Rosenberg, J. Francis, C.Z. Zhang, O. Cohen, C. Oh, H. Ding, M. Lloyd, S. Mahmud, K. Helvie, M.S. Merrill, R.S. Santiago, E. O'Connor, S.H. Jeong, R. Leeson, R. Barry, J. Kramkowski, Z. Zhang, L. Polacek, J.G. Lohr, M. Schleicher, E. Lipscomb, A. Saltzman, N.M. Oliver, L. Marini, L. Harshman, S. Tolaney, E. Van Allen, E.P. Winer, N.U. Lin, M. Nakabayashi, M.E. Taplin, C.M. Johannessen, L. Garraway, T.R. Golub, J.S. Boehm, N. Wagle, G. Getz, J.C. Love, M. Meyerson. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications*, 2017. PubMed Central PMCID: PMC5673918

5. Genomic and phenotypic profiling of single cells stands to uncover the molecular underpinnings of cancer. We have developed novel approaches to isolate, amplify, sequence, and analyze whole-genomes and whole-exomes of single cells such as circulating tumor cells. We have found that CTCs harbor most early and metastatic trunk mutations found in a patient's cancer. We have also developed methods to characterize whole-genome amplification of single cells and to study protein secretion at the single-cell level. For instance, using a nanowell-based approach to compartmentalize single cells, we profiled the secretions of ELR+ CXC chemokines from thousands of single tumor and normal cells of human primary colorectal tumors. We showed that both tumor and normal cells secrete ELR+ CXC chemokines but that a small fraction of single cells secrete at any given instance (~1 hr intervals). In addition, we demonstrated that secretory phenotypes of single tumor and normal cells can evolve and are often discordant with phenotypes observed in bulk.

- J.G. Lohr*, **V.A. Adalsteinsson***, K. Cibulskis*, A.D. Choudhury, M. Rosenberg, P. Cruz-Gordillo, J. Francis, C.Z. Zhang, A.K. Shalek, R. Satija, J.T. Trombetta, D. Lu, N. Tallapragada, N. Tahirova, S. Kim, B. Blumenstiel, C. Sougnez, A. Lowe, B. Wong, D. Auclair, E.M. Van Allen, M. Nakabayashi, R.T. Lis, G.S.M. Lee, T. Li, M.S. Chabot, A. Ly, M.E. Taplin, T.E. Clancy, M. Loda, A. Regev, M. Meyerson, W.C. Hahn, P.W. Kantoff, T.R. Golub, G. Getz, J.S. Boehm, J.C. Love, Whole exome sequencing of CTCs as a window into metastatic prostate cancer, *Nature Biotechnology*, 2014. PubMed Central PMCID: PMC4034575
- C.Z. Zhang*, **V.A. Adalsteinsson***, J. Francis, J. Jung, C. Maire, H. Cornils, M. Leibowitz, D. Pellman, K.L. Ligon, M. Meyerson, J.C. Love, Calibrating genomic and allelic coverage bias in single-cell sequencing, *Nature Communications*, 2015. PubMed Central PMCID: PMC4922254
- **V.A. Adalsteinsson**, N. Tahirova, N. Tallapragada, X. Yao, L. Champion, A. Angelini, T.B. Douce, C. Huang, B. Bowman, C.A. Williamson, D.S. Kwon, K.D. Wittrup, J.C. Love, Single cells from human primary colorectal tumors exhibit polyfunctional heterogeneity in secretions of ELR+ CXC chemokines, *Integrative Biology*, 2013. PubMed Central PMCID: PMC4034677

Complete List of Published Work in My Bibliography:

<http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/49168493>