

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ILLUMINA INC.,
Petitioner,

v.

MOLECULAR LOOP BIOSCIENCES, INC.,
Patent Owner.

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

REPLY DECLARATION OF PAUL T. SPELLMAN, PH.D.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. MATERIALS CONSIDERED	1
III. LEVEL OF ORDINARY SKILL IN THE ART	2
IV. CLAIM CONSTRUCTION	3
A. Limiting “Template Cross-Over Error” and “Cross-Over Error” to Chimeric Molecules “Created During Cluster Amplification” Is Improper	3
B. Adding “Error” Is Improper	15
V. CLAIMS 1-8 OF THE ’852 PATENT WOULD HAVE BEEN OBVIOUS.....	16
A. Ground 1: Claims 1-8 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Gloor</i>	17
1. Dr. Adalsteinsson’s Characterization of the State of the Art Is Inaccurate and Incomplete.....	17
2. Dr. Adalsteinsson Fails to Read <i>Parameswaran</i> from the Perspective of a POSA in December 2010	34
3. Dr. Adalsteinsson’s Analysis of the “Discarding” Step Ignores Important Disclosures in <i>Parameswaran</i> and <i>Gloor</i>	46
4. Dr. Adalsteinsson’s Reasons for <i>Not</i> Combining <i>Parameswaran</i> and <i>Gloor</i> Lack Scientific Merit	54
5. Dr. Adalsteinsson’s Focus on Absolute Number of Barcodes Ignores Context.....	62
6. Dr. Adalsteinsson’s Argument on Reasonable Expectation of Success Is Flawed.....	65

B.	Ground 2: Claims 1-8 Would Have Been Obvious Over <i>Parameswaran and Bentley</i>	68
C.	Claims 1-8 Would Have Been Obvious Under Either My Construction or Dr. Adalsteinsson's Construction	70
VI.	CONCLUSION.....	74

I. INTRODUCTION

1. I, Dr. Paul T. Spellman, have been retained as an independent expert in the field of genetics and DNA sequencing. I previously provided a declaration dated July 1, 2024 (EX1003) on behalf of Petitioner Illumina, Inc., in the above-captioned *inter partes* review (“IPR”) of U.S. Patent No. 11,041,852 (“the ’852 patent”; EX1001). My background and qualifications are set forth in my previous declaration. I now provide my expert opinions in response to arguments made by Dr. Viktor Adalsteinsson in his declaration (EX2001).

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

II. MATERIALS CONSIDERED

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

4. In addition to the materials I considered in forming my opinions expressed in my previous declaration, I have considered additional materials cited in this declaration, which are listed in **Appendix A**. I have also considered the declaration of Dr. Adalsteinsson and materials cited by him. I reserve the right to rely on documents cited in the appendices to this declaration, and to supplement my opinions in view of new materials and information that become available to me during this proceeding.

III. LEVEL OF ORDINARY SKILL IN THE ART

5. In my previous declaration, I opined that a person of ordinary skill in the art (“POSA”) here 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. EX1003, ¶20. I also explained my understanding that an assessment of a patent’s claims should be undertaken from the perspective of a POSA as of the effective filing date of the patent claims, which I was asked to assume to be December 23, 2010. *Id.*, ¶18.

6. Dr. Adalsteinsson does not disagree with my definition of a POSA or my understanding that the claims of the ’852 patent should be evaluated from the perspective of a POSA as of December 23, 2010. *See* EX2001, ¶¶19-22, 50.

IV. CLAIM CONSTRUCTION

7. In my previous declaration, I discussed that the term “crossover”—which appears four times in the specification of the ’852 patent—is used to describe the formation of chimeric molecules that contain a combination of two template nucleic acids derived from different samples. EX1003, ¶122. Based on this, I adopted and applied the meaning of the term “template crossover error” as used in the ’852 patent: sequences of chimeric molecules made from template nucleic acids derived from different samples. *Id.*

8. In response, Dr. Adalsteinsson offers an alternative construction of the terms “template cross-over error” and “cross-over error”: error from sequences of chimeric molecules created during cluster amplifications and made from template nucleic acids derived from different samples. EX2001, ¶¶24-30. I underlined the words that Dr. Adalsteinsson added to my construction. As discussed below, it is my opinion that these additional words should not be added to the construction of the terms “template cross-over error” and “cross-over error.”

A. Limiting “Template Cross-Over Error” and “Cross-Over Error” to Chimeric Molecules “Created During Cluster Amplification” Is Improper

9. Dr. Adalsteinsson agrees that the claim terms “template cross-over error” and “cross-over error” refer to chimeric molecules made from template nucleic acids derived from different samples. But he is incorrect that these terms

should be limited to chimeric molecules “created during cluster amplification” because this contradicts the claim language, the specification, the prosecution history of the ’852 patent, and the general knowledge in the art.

10. Claim Language: Dr. Adalsteinsson proposes to limit the terms “template cross-over error” and “cross-over error” to chimeric molecules “created during cluster amplification.” But he is silent on how his proposal aligns with the text of claim 1 of the ’852 patent. See EX2001, ¶¶24, 26-30. As discussed below, the claim language demonstrates that his proposal is incorrect.

11. As reproduced below and highlighted in red, claim 1 of the ’852 patent recites a method for reducing “template cross-over error *introduced during sequencing workflow*” (emphasis added). Cluster amplification is just one of many steps that occurs during the sequencing workflow. It is also just one of many steps when chimeric molecules made from template nucleic acids derived from different samples can arise. Since the claim does not identify any specific steps during the sequencing workflow where “template cross-over error” or “cross-over error” must be introduced, a POSA would have understood that the terms “template cross-over error” and “cross-over error” encompass chimeric molecules created at *any* stage during the sequencing workflow, not just “created during cluster amplification,” as Dr. Adalsteinsson suggests.

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.

12. Other language in claim 1 further supports that claim 1 is directed to reducing cross-over error at any stage of the sequencing workflow. For example, claim 1 recites various steps during the sequencing workflow, including the step of “amplifying” nucleic acid templates to “form clusters” (highlighted in blue above).

Had the inventors intended for the terms “template cross-over error” and “cross-over error” to cover only chimeric molecules “created during cluster amplification,” they would have included that limitation.

13. Instead, it is notable that this “amplifying” step requires only that at least one of the clusters “*comprises* a chimeric sequence.” A POSA reading this step would not have understood that a chimeric sequence must be “created” during cluster amplification. Instead, a POSA would have understood it to mean that the chimeric sequence could be created before or during cluster formation, so long as at least one of the clusters contains the chimeric sequence.

14. This is consistent with Dr. Adalsteinsson’s deposition testimony regarding the “amplifying” step. Dr. Adalsteinsson acknowledged that “the chimeric sequence” in this step “*includes* a type of chimera associated with template cross-over error.” EX1055 (Adalsteinsson Tr.), 88:18-90:12 (emphasis added). Dr. Adalsteinsson reiterated that “it’s *not limited* to” cross-over error arising on the surface of a flow cell because “[o]ther types of chimeras can arise in the sequencing process.” *Id.* (emphasis added); *see also id.*, 91:14-92:5 (Dr. Adalsteinsson explaining: “It’s possible that other chimeras can be formed and sequenced.”). The plain text of claim 1 of the ’852 patent supports my claim construction that the terms “template cross-over error” and “cross-over error” are sequences of chimeric

molecules made from template nucleic acids derived from different samples, regardless of when and where in the sequencing workflow they arise.

15. Specification: Ignoring the plain text of claim 1 of the '852 patent, Dr. Adalsteinsson focuses on select statements in the specification. EX2001, ¶¶26-29. Specifically, Dr. Adalsteinsson contends that “the specification says that chimeric molecules are formed after ‘the amplification process begins’ when two template molecules overlap.” *Id.*, ¶27. He then contends that 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.” *Id.* He further contends that, because the '852 patent “provides a solution” to this specific type of chimeric molecules, the claims therefore must be limited to reducing error arising from those chimeric molecules. *Id.*, ¶28. I disagree.

16. Dr. Adalsteinsson ignores express statements in the specification of the '852 patent emphasizing that the creation of chimeric molecules is not “limited by any particular theory or mechanism.” EX1001, 6:25-29; *see also id.*, 6:64-7:1 (“Without being limited by any particular theory or mechanism of action, . . . the nucleic acid from sample A and the nucleic acid from sample B landed within a cluster radius on the solid support.”). Dr. Adalsteinsson also ignores the specification’s statement that its dual-indexing method is “not intended to be limited to use of the sequencing method outlined above, as essentially any sequencing

methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used.” *Id.*, 20:37-46. These express statements from the ’852 patent specification support a construction that is not limited to detecting chimeras “created” during cluster amplification on Illumina flow cells.

17. Prosecution History: I also disagree with Dr. Adalsteinsson’s characterization of the ’852 patent’s prosecution history. EX2001, ¶30. During prosecution, Applicant distinguished cluster formation on an Illumina flow cell versus cluster formation on the beads used in Roche 454 sequencing. EX1007, 304-306 (“Parameswaran does not mention forming clusters in a flow cell. . . . In pyrosequencing, one prepares a series of bead-bound templates . . .”). Applicant did not use the term “template cross-over error” or “cross-over error.” *Id.* Nor did Applicant identify any specific step where chimeric molecules would arise. *Id.* Likewise, the Examiner understood Applicant’s arguments to be about cluster formation on a flow cell versus on beads, noting that the “claims should be amended to read on Illumina-type clusters” since “bead emulsion amplification of Parameswaran et al. forms clusters on beads.” *Id.*, 311.

18. Despite this, Dr. Adalsteinsson contends that “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,” and that this amendment is consistent with his reading of “cross-over error” as occurring

during amplification. EX2001, ¶30. But as reproduced below, this amendment says nothing about when and where chimeric sequences were first created.¹ EX1007, 302.

47. (Currently amended) 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.

19. This amendment merely requires amplifying nucleic acid templates on a surface of a flow cell (as opposed to, e.g., beads in *Parameswaran*), and says nothing about when or how the chimeric sequence arises. As illustrated below, the clusters on the surface of the flow cell would “comprise[] a chimeric sequence” regardless of whether the chimeric molecules were created before or during cluster amplification. *See infra*, ¶¶20-25. Contrary to Dr. Adalsteinsson’s characterization,

¹ I understand that the underlined portion of claim 47 is the amendment at issue, adding the phrase “on a surface of a flow cell” to the claim.

the prosecution history of the '852 patent does not support his proposal that the claims “template cross-over error” and “cross-over error” should override the plain language of the claims to be limited to chimeric molecules “created during cluster amplification.”

20. General Knowledge in the Art: Dr. Adalsteinsson’s narrower reading of the terms “template cross-over error” and “cross-over error” is also inconsistent with the general knowledge in the art about the formation of chimeric sequences during the sequencing workflow.

21. As discussed in my previous declaration, it was well known by December 2010 that chimeric molecules could form at different stages of the sequencing workflow, including cluster amplification. *See* EX1003, ¶¶67-69. This makes sense because chimeric molecules can form whenever and wherever polymerase chain reaction (“PCR”) amplification is performed during the sequencing workflow, regardless of any specific step and any specific location. *See* EX1017 (Lahr), 1 (“Chimeras are formed when incompletely extended DNA fragments anneal to closely related sequences generating recombinants between starting templates.”); EX1060 (Gorzer), 1 (“When PCR amplification is carried out using a mixture of two or more non-identical but similar genotype target sequences as templates, this can lead potentially to the formation of artificial chimeras between the amplification products.”).

22. As of December 2010, sequencing workflows in next-generation sequencing (“NGS”) platforms often involved several PCR amplification steps, including PCR performed after pooling indexed libraries from multiple samples but before cluster amplification. For example, as Dr. Adalsteinsson acknowledged, sequencer-specific adaptors were commonly attached to templates in pooled samples via PCR amplification, including for Illumina sequencing. *See* EX2001, ¶35; EX1005 (Gloor), 23-24 (sending pooled libraries for Illumina sequencing at a core facility); EX1055 (Adalsteinsson Tr.), 75:17-76:3. Similarly, templates in pooled samples were often amplified before being loaded onto NGS sequencers. *See, e.g.*, EX1015 (Meyer 2010), 9 (“amplifying the library pool” using primers IS5 and IS6 and then sequencing the amplified products using an Illumina sequencer); EX1021 (Craig), 15-16 (“pooled” libraries were “PCR-enriched using Phusion DNA polymerase PCR mix and adapter compatible primers 1.1 and 2.1 (Illumina)” and “sequenced on an Illumina GA”); EX1055 (Adalsteinsson Tr.), 44:1-12 (“Q. So do you have an understanding as to whether or not in December 2010 the standard Illumina sequencing workflow included a PCR step to selectively enrich DNA fragment having adapter molecules on both ends? A. If the library’s amplified before loading onto the flow-cell, that will be done.”). A POSA would have known that these PCR amplifications occurring before cluster amplification could create chimeric sequences. *See* EX1015 (Meyer 2010), 9 (“To avoid cross-contamination

of samples through jumping PCR (Meyerhans et al. 1990 [EX1018]), pools of indexed libraries should be amplified with a minimum number of PCR cycles”); *see infra* § V.A.1.

23. Notably, as illustrated below, sequences from chimeric molecules formed *before* cluster amplification (e.g., during PCR enrichment with pooled samples; **Figure 1**) would contain mismatched barcodes, as would chimeric molecules formed on the flow cell during cluster amplification (EX2001, ¶27; **Figure 2**).

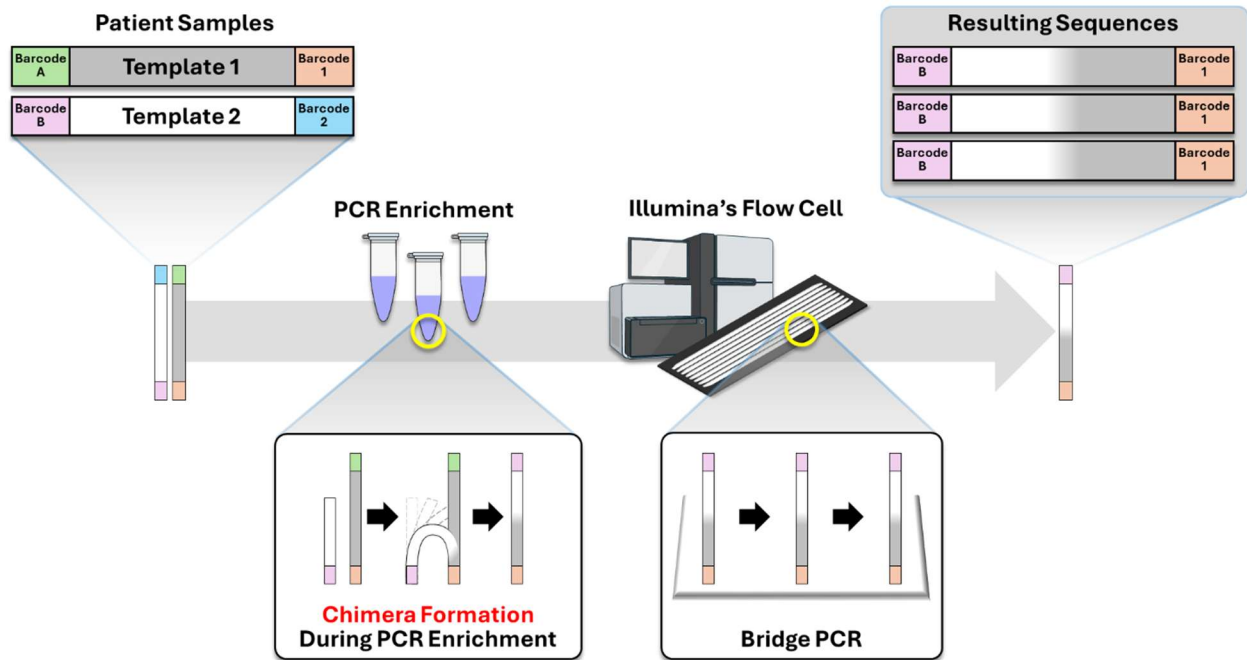


Figure 1. Sequences of Chimeric Molecules Created Before “Cluster Amplification”

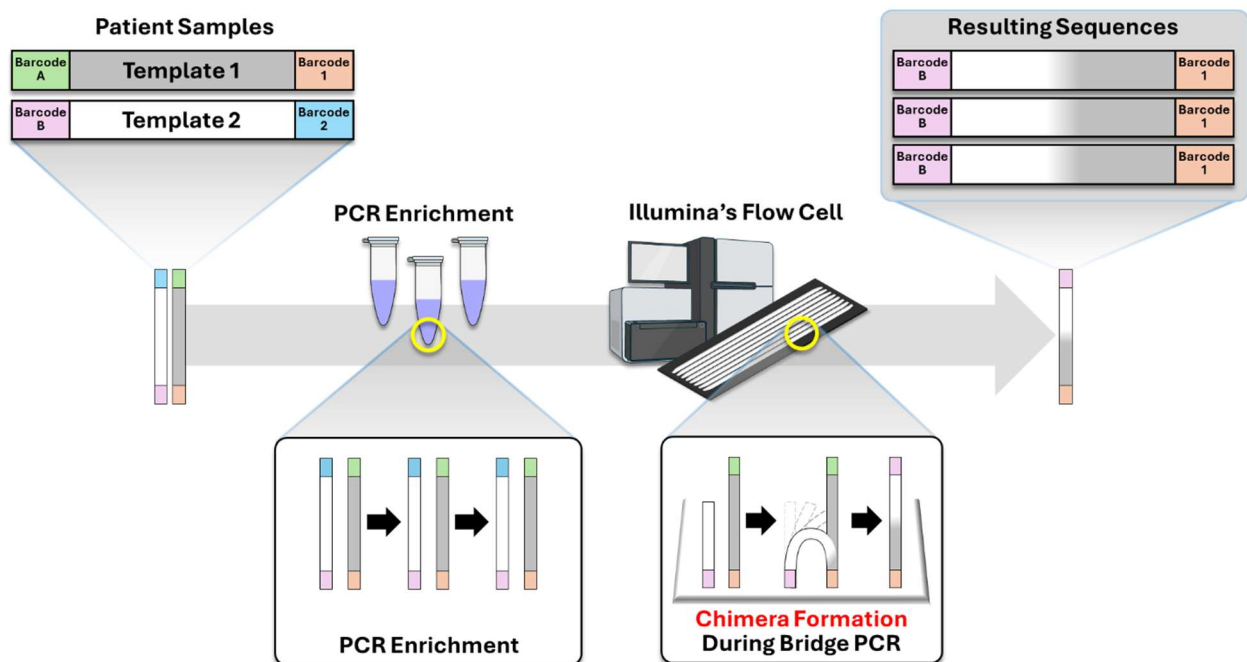


Figure 2. Sequences of Chimeric Molecules Created “During Cluster Amplification”

24. In the context of the '852 patent claims, sequences of chimeric molecules made of templates from different samples “created during cluster amplification” are indistinguishable from those created at other stages of the sequencing workflow—both contain pairs of mismatched identifiers that should not be found together.² The existence of such sequences constitutes “cross-over error” regardless of when or how those sequences arise. Further, discarding sequences with mismatched identifiers would discard all of these chimeric sequences, regardless of when and how they formed during the sequencing workflow. In other words, the '852 patent does not teach how to selectively reduce cross-over errors “created during cluster amplification” instead of cross-over errors created at other points during the sequencing workflow, further indicating that Dr. Adalsteinsson’s proposed construction is incorrect.

25. Consistent with this, the specification provides no guidance on how to selectively reduce sequences from chimeric molecules made of templates from

² Similarly, sequences from these chimeric molecules are indistinguishable from errors created from “mixed clusters” containing multiple template nucleic acids. *See, e.g.*, EX1006 (Bentley), 18. Neither the claims nor the specification explains how a POSA could selectively reduce cross-over errors “created during cluster amplification” versus these errors due to “mixed clusters.”

different samples “created during cluster amplification” using identifier sequences alone. *See generally* EX1001. Likewise, the claims of the ’852 patent recite no specific step for doing so. *See id.*, cls. 1-8. Given that chimeric molecules made of templates from different samples could form at any stage during the sequencing workflow and that both the specification and the claims are silent on how to selectively reduce sequences from the chimeric molecules created at a particular stage during sequencing workflow, a POSA would have understood that the claimed methods are not limited to reducing sequences from chimeric molecules made of templates from different samples “created during cluster amplification.”

26. For the reasons discussed above and in my prior declaration, the terms “template cross-over error” and “cross-over error” should be construed as: sequences of chimeric molecules made from template nucleic acids derived from different samples, without narrowing the claim language to cross-over error arising during cluster amplification. EX1003, ¶122.

B. Adding “Error” Is Improper

27. Dr. Adalsteinsson does not explain why he added the word “error” in his proposed construction.

28. Claim 1 recites “reducing cross-over error introduced during sequencing workflow.” My proposed construction explains what that error is—the existence of “sequences of chimeric molecules made from template nucleic acids

derived from different samples.” Merely repeating the claim term “error” adds nothing to the construction.

29. Dr. Adalsteinsson appears to suggest that my construction is somehow inconsistent with the last clause of claim 1 and/or requires reducing the formation of sequences from chimeric molecules. *See* EX2001, ¶25 (“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”). Neither is correct. The last clause of claim 1 recites “discarding sequence reads . . . that contain said combination of identifier sequences that are different than any of said distinct pairs.” EX1001, cl. 1. The clause does not state that sequence reads “obtained from chimeric molecules” are discarded. Further, by discarding sequences containing mismatched identifier sequences (and not by “stopping” the formation of chimeric sequences), the claimed methods reduce “sequences of chimeric molecules made from template nucleic acids derived from different samples.” Thus, my construction is consistent with the claim language. Dr. Adalsteinsson’s proposed addition of a superfluous word into the construction, which itself is one of the words being construed, is irrelevant.

V. CLAIMS 1-8 OF THE ’852 PATENT WOULD HAVE BEEN OBVIOUS

30. Dr. Adalsteinsson does not suggest that the combination of *Parameswaran* (unique dual-indexing) with either *Gloor* or *Bentley* (both disclosing

Illumina sequencing) is incapable of identifying sequences of chimeric molecules made of templates from different samples formed during sequencing workflow, including those formed during cluster amplification. *See generally* EX2001. Instead, Dr. Adalsteinsson argues that a POSA would not have been motivated, or had a reasonable expectation of success, to combine these references because the claimed methods (under his construction) address a problem that he contends was unknown in the art as of December 2010. *See* EX2001, §§XI, XII. I disagree with his opinion, and address each of his arguments below.

A. Ground 1: Claims 1-8 Would Have Been Obvious Over *Parameswaran* and *Gloor*

1. Dr. Adalsteinsson’s Characterization of the State of the Art Is Inaccurate and Incomplete

31. In my previous declaration, I explained that the formation of chimeric molecules was a known problem for NGS platforms as of December 2010. EX1003, ¶¶67-69. Dr. Adalsteinsson disagrees, contending that there was a “general lack of awareness of cross-over error in NGS machines” as of December 2010. EX2001, ¶¶105-116; *see also id.*, ¶¶49, 128-129, 136, 150. I respectfully disagree.

a. Chimeric Molecules, Including Those “Created During Cluster Amplification,” Were a Known Problem

32. Contrary to Dr. Adalsteinsson’s characterization, and as explained in my prior declaration, chimeric molecules, including those “created during cluster

amplification,” were a known problem for NGS platforms such as Illumina sequencers as of December 2010.

33. Chimeric Molecules and NGS Platforms: By December 2010, it was well known that PCR amplification would give rise to chimeric molecules formed from multiple template nucleic acids. EX1003, ¶67. These chimeric molecules are artificial combinations of multiple individual nucleic acids that are not present in the starting template nucleic acids. *See, e.g.*, EX1018 (Meyerhans), 1 (“[R]ecombinant or ‘shuffled’ molecules, i.e. molecules composed of parts of ***two different sequences artificially combined*** during PCR may be produced.”) (emphasis added). This phenomenon was fundamental to PCR amplification involving different template nucleic acids. EX1063 (Smyth), 1. For example, incomplete primer extension during the extension phase of PCR amplification was understood to create partially extended products. These products could bind other, nearby templates, act as primers for the next round of extension, and create artificial chimeras as a result. *Id.* This type of error is typically referred to as “mispriming.” EX1003, ¶¶23, 59.

34. A POSA would have known as of December 2010 that the concern of artificial chimeric molecules was relevant for NGS platforms, which relied on PCR amplification at multiple stages during the sequencing workflow. EX1003, ¶¶60-61. For example, PCR amplification was commonly used to generate DNA libraries for individual samples. *See* EX1004 (Parmeswaran), 3-4. PCR amplification was

commonly used to amplify DNA libraries pooled from multiple samples. *See* EX1015 (Meyer 2010), 9. PCR amplification was also commonly used to attach sequencer-specific adaptors. EX1055 (Adalsteinsson Tr.), 43:1-6, 75:17-76:3. These DNA libraries (from individual or pooled samples) were commonly immobilized onto beads (Roche 454) or a flow cell (Illumina), and another PCR amplification was commonly used to generate millions of copies for each fragment in the libraries to produce sufficient signal to be detected during the sequencing step. EX1003, ¶¶60-61.

35. Any of these PCR amplification steps in the sequencing workflow could create chimeric molecules. Because of this, by December 2010, NGS platforms, including Roche 454 and Illumina sequencers alike, were recognized as an “ideal ‘breeding ground’” for chimeric molecules. EX1034 (Creer), 10; *see id.* (noting that NGS platforms “are based on the amplification of homologous regions from a large number of potentially highly related organisms,” which requires “empirical work . . . to assess the level of chimera formation in community-based PCR”); EX1033 (Dai), 8 (“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.”).

36. Chimeric Molecules and Multiplex Sequencing: By December 2010, it was well known that multiplex sequencing is particularly vulnerable to the formation

of chimeric molecules. For multiplex sequencing, it was common to combine closely related template nucleic acids from multiple samples. In *Gloor*, for example, the V6 regions of bacterial 16S rRNA were sequenced from 272 clinical samples. *See, e.g.*, EX1005 (*Gloor*), Abstract, 2. These regions were amplified using common primers, indicating high sequence similarities among the target nucleic acids of interest. *Id.*, 4-5. High sequence similarities among template nucleic acids were problematic because the generation of chimeric molecules was known to increase in frequency when template nucleic acids are similar to one another. *See, e.g.*, EX1017 (*Lahr*), 1 (“Chimeras are formed when incompletely extended DNA fragments anneal to ***closely related sequences*** generating recombinants between starting templates.”) (emphasis added); EX1060 (*Gorzer*), 1 (“When PCR amplification is carried out using a mixture of two or more ***non-identical but similar genotype target sequences*** as templates, this can lead potentially to the formation of artificial chimera between the amplification products.”) (emphasis added); EX1063 (*Smyth*), 2 (“Unfortunately, ***highly similar sequences are the most likely to form chimeras during PCR***, meaning the most commonly formed chimeras are the most difficult to computationally detect.”) (emphasis added). Physically, this is caused by regions of high sequence identity increasing the chance of incomplete extension products being able to bind to and serve as primers to neighboring templates.

37. Chimeric molecules were a known type of error that formed during multiplex sequencing. *Gloor*, for example, addressed the issue of chimeric sequences formed during multiplex sequencing using an Illumina sequencer. EX1005 (*Gloor*), 8, 22. *Gloor* specifically noted that these chimeric sequences likely arose “*during PCR* or because of *overlapping spots on the solid support* when imaged during DNA sequencing.” *Id.*, 8 (emphasis added). A POSA reading *Gloor* would have understood that these sequences included sequences from chimeric molecules formed between different templates that arose at various stages during multiplex sequencing, including after sample pooling for Illumina-specific adaptor attachment and from mixed clusters on Illumina’s flow cells. *See id.*; EX1055 (*Adalsteinsson Tr.*), 44:1-12, 75:17-76:3.

38. *Meyer 2010* expressly warned that chimeric molecules can be formed between templates derived from different samples in multiplex sequencing. In *Meyer 2010*, the authors created barcoded libraries from different samples. EX1015 (*Meyer 2010*), 1. The authors attached barcodes to templates from each sample to create a barcoded library. *Id.*, 8-9. The authors then “pool[ed] the indexed libraries in equimolar ratios.” *Id.*, 9. The pooled libraries were further amplified and then sequenced using an Illumina sequencer. *Id.* Notably, for the steps of amplifying the pooled indexed libraries and sequencing, the authors warned that “jumping PCR” could occur during PCR amplification of barcoded libraries. *Id.*

39. When chimeric molecules are formed between template nucleic acids from different samples, a sequence associated with one sample could be misassigned to a different sample, thereby creating false positives and false negatives amongst multiple samples. EX1060 (Gorzer), 1 (“When PCR amplification is carried out using a mixture of two or more non-identical but similar genotype target sequences as templates, this can lead potentially to the formation of artificial chimeras between the amplification products.”). The consequences of such errors could be significant in many contexts in which DNA sequencing is used, as detailed below. *See infra* §§ V.A.2-5.

40. Chimeric Molecules and Illumina Sequencers: As of December 2010, it was well known that the Illumina sequencing workflow generated chimeric sequences, which were problematic for applications that demanded accurate sample assignment and sequencing.

41. *Campbell*, for example, sequenced two cancer cell lines to identify cancer-causing chimeras, which are fused-nucleic acids that exist in nature, as opposed to those created during the sequencing workflow. EX1057 (Campbell), 2. In that process, *Campbell* identified a large number of chimeric sequences generated during Illumina sequencing workflow. From the two cell lines, *Campbell* “generated 36.2 million paired reads of 29-36 bases at each end.” *Id.* About “18.8 million (52%)

reads were of sufficient quality that both ends mapped back to the current reference genome assembly.” *Id.*

42. Among those 18.8 million reads, about “1.2 million reads” “suggested a possible rearrangement”—i.e., chimeric sequences. EX1057 (Campbell), 2. Notably, *Campbell* observed that many of these sequences arose “presumably due to incorrect mapping induced by sequencing errors.” *Id.* Even among the chimeric sequences that did not indicate sequencing errors (7,042 read pairs from the first cell line and 19,071 from the second cell line), *Campbell* was able to validate only about 400 read pairs as biological chimeras (325 read pairs from the first cell line and 84 from the second cell line). *Id.* *Campbell* demonstrates that the rate of biological chimeric sequences was low (~26,000 among 18.8 million sequences) while the amount of artificial chimeric sequences generated during the Illumina sequencing workflow was high (nearly 1.2 million chimeric sequences).

43. *Campbell* reported that chimeric sequences excluded from its analysis included PCR-induced artifacts generated on the Illumina flow cell. Specifically, *Campbell* excluded sequence reads “where the two ends mapped back to within 500 bp of one another, but with one of the two ends in the incorrect orientation,” as “these were likely to be artifacts resulting from either mispriming of the Solexa [Illumina] sequencing oligonucleotide ***within the PCR colony*** or intramolecular rearrangements generated during library amplification.” EX1057 (Campbell), 7

(emphasis added); EX1010 (Kircher), 5 (“Solexa sequencing” is “now called the Illumina Genome Analyzer”).

44. Studies after *Campbell* continued to observe PCR-induced artifacts created during cluster amplification on an Illumina sequencer. For example, *Varela* used an Illumina sequencer to analyze breast cancer samples to identify somatic mutations. EX1066 (*Varela*), 2-3. But *Varela* detected artifacts during that process, which *Varela* attributed to “mispriming of the sequencing oligonucleotide ***within the bridge amplified cluster.***” *Id.*, 10 (emphasis added). As in *Campbell*, *Varela* excluded these artifacts from its downstream analysis. *See id.*

45. *Maher* is another example illustrating the propensity of Illumina sequencers to generate chimeric sequences. *Maher* used an Illumina sequencer to detect biologically relevant chimeras associated with cancer, i.e., chimeric molecules that existed before sequencing. EX1061 (*Maher*), 1. Because it was important to differentiate true chimeras from artificial chimeras created during the sequencing workflow, *Maher* implemented multiple filtering steps. For instance, *Maher* imposed a requirement that “[a]t least 2 independent mate pairs are required to support a chimera nomination.” *Id.*, 6. During the filtering steps, *Maher* also detected candidate chimeras that had “identical mate pair reads that were in close proximity on the flow cell,” leading to “overrepresentation of a subset of chimeras.” *Id.*, 8. As *Maher* explained, a POSA would have understood that these candidate

chimeras were an “artifact of the analysis pipeline” (i.e., were created during the sequencing workflow) and should be excluded to ensure sequencing accuracy. *Id.* After imposing these rigorous filtering steps, *Maher* identified 31 “chimera candidates” (*see id.*, 17 (Table S2)) from human brain reference (“HBR”) total RNA among 114,084 potential chimeric sequences (*see id.*, 16 (Table S1)). This again confirms the known problem of artificial chimeric sequences created during the Illumina sequencing workflow.

46. A POSA would have understood that many of these chimeric sequences arose from chimeric molecules created during cluster amplification on Illumina’s flow cells. As noted above, it was well known that Illumina sequencers used PCR amplification, sometimes called bridge PCR or bridge amplification, to form “clusters” for sequencing on the surface of a flow cell. EX1065 (Taylor), 5 (the Illumina platform uses a “PCR process known as bridge PCR to achieve clonal amplification of sequencing templates on the surface of a glass flow cell”); EX1012 (Shendure), 4 (in the Illumina platform, “[a]mplified sequencing features are generated by bridge PCR”).

47. During his deposition, Dr. Adalsteinsson suggested that bridge PCR is qualitatively different from other forms of PCR because it uses “isothermal denaturation.” EX1055 (Adalsteinsson Tr.), 41:19-42:5. But the purpose and result of “isothermal denaturation” are the same as “thermal denaturation” used in typical

PCR. The “isothermal denaturation” simply means that the denaturation steps during PCR are conducted at a constant temperature and use chemical means (e.g., formamide) to separate nucleic acid strands, whereas “thermal denaturation” separates the strands by increasing the temperature. These differences do not change the fundamental characteristics of PCR that lead to chimera formation. In December 2010, it was known that the extension phase of PCR, not the denaturation phase, was principally responsible for chimera formation. EX1020 (Odelberg), 1 (“recombinants can be generated during a single round of primer extension in the absence of subsequent heat denaturation”). Bridge amplification, like traditional PCR, involves polymerases extending template nucleic acids, which is the fundamental step leading to incomplete extension and the formation of chimeric molecules between nearby template nucleic acids. EX1011 (Metzker), 3-4.

48. The “incomplete extension of the template” on the surface of a flow cell would have been particularly problematic when different templates were seeded nearby on the flow cell. Indeed, as Dr. Adalsteinsson acknowledged and the prior art confirms, for Illumina platforms, it was known that “a single template nucleic acid is intended to be clonally amplified to form a cluster” but “at times, more than one template nucleic acid molecule may seed each” cluster amplification. *See* EX2001, ¶38; EX1067 (Whiteford), 3 (“In the ideal scenario, discrete clusters are grown, each from a single DNA template. However, mixed clusters can be grown

starting from more than one template, in which case the base calls of the cluster cannot be accurately determined.”). Incompletely extended nearby templates were known to create an ideal opportunity for the formation of chimeric molecules. *See* EX1063 (Smyth), 1 (“the majority of chimeras are thought to arise due to incomplete primer extension during the extension phase of the PCR”). As illustrated in the multiple studies discussed above, a POSA in December 2010 would have understood that these PCR-mediated events would result in chimeric molecules formed between different templates on Illumina’s flow cells.

b. Dr. Adalsteinsson Wrongly Dismisses Relevant Prior-Art Studies

49. In my previous declaration, I cited exemplary references explaining the mechanism behind the formation of chimeric molecules (including EX1017 (Lahr), EX1018 (Meyerhans), EX1019 (Paabo), EX1020 (Odelberg)) as well as references expressly characterizing chimeric molecules as a problem for NGS platforms (including EX1033 (Dai) and EX1034 (Creer)). Dr. Adalsteinsson disregards these references, because they purportedly did not show that “PCR recombination of tagged nucleic acids *from different samples* were known to arise *and produce chimeras with cross-overs of sequence tags.*” EX2001, ¶106 (emphasis in original); *see id.*, ¶¶105-113 (Dr. Adalsteinsson discussing EX1017 (Lahr), EX1018 (Meyerhans), EX1019 (Paabo), EX1020 (Odelberg), EX1033 (Dai), and EX1034 (Creer)).

50. Dr. Adalsteinsson fails to read these references from the perspective of a POSA in December 2010. As discussed above, references such as EX1017 (Lahr), EX1018 (Meyerhans), EX1019 (Paabo), and EX1020 (Odelberg) provided important insights into the mechanism of chimera formation during PCR amplification, a process that was fundamentally consistent in different contexts despite technological advances in sequencing platforms. EX1003, ¶67. A POSA in December 2010 would have understood from these foundational references that chimeric molecules would result during PCR amplification whenever multiple DNA templates were co-amplified.

51. A POSA in December 2010 also would have understood that PCR recombination occurred between tagged template nucleic acids “*from different samples*.” As discussed above, by December 2010, this had already been reported in the literature. *Meyer 2010* is one example where the authors expressly warned that “jumping PCR” could occur during PCR amplification of barcoded libraries. EX1015 (Meyer 2010), 9. *Gloor* is another example, reporting that chimeric sequences among *tagged* nucleic acids occurred during *multiplex* sequencing using an Illumina sequencer. EX1005 (Gloor), 8 (citing EX1017 (Lahr) and noting that “[c]himeric sequences can arise during PCR or because of overlapping spots on the solid support”), 23-24 (“the PCR products were mixed together to give the final sample sent for Illumina sequencing”); EX1055 (Adalsteinsson Tr.), 75:17-76:3.

52. EX2004, cited by Dr. Adalsteinsson, relied on multiple references that I cited in my earlier declaration to explain chimera formation in the context of Illumina's multiplex sequencing workflow. *See* EX2001, ¶¶43, 49, 136, 150 (citing EX2004). Specifically, EX2004 explained that "it is usually inevitable to amplify libraries after capture" before Illumina sequencing. EX2004, 2. EX2004 then cited EX1017 (Lahr), EX1018 (Meyerhans), EX1019 (Paabo), and EX1020 (Odelberg), stating that "*PCR can produce chimeras by recombining different templates molecules* (a process often referred to as 'jumping PCR')," which "introduce significant levels of *sample cross-contamination*." EX2004, 2 (emphasis added). Thus, Dr. Adalsteinsson's cited evidence confirms that a POSA in December 2010 would have appreciated that "PCR recombination of tagged nucleic acids *from different samples*" could occur during sequencing workflow and would "*produce chimeras with cross-overs of sequence tags*." *See* EX2001, ¶106 (emphasis in original).

53. Dr. Adalsteinsson cites additional articles that he contends show that cross-over error was not a recognized problem for NGS platforms such as Illumina sequencers as of December 2010. EX2001, ¶¶49, 114, 128-129, 136, 150 (citing EXS2003-2007). Again, I disagree with Dr. Adalsteinsson's interpretation of this evidence.

54. EX2005: Citing EX2005, Dr. Adalsteinsson states that “literature at the time of the invention explained that cross-over error for flow-cell based NGS was not thought to be a problem.” EX2001, ¶114. But as Dr. Adalsteinsson clarified during his deposition, the use of the term “cross-over error” in EX2005 was *not* a reference to “what the patent discloses, which is template cross-over error that occurs during cluster amplification on the surface of a flow cell.” EX1055 (Adalsteinsson Tr.), 25:22-26:15. Instead, EX2005 “examine[d] whether sequences from one library could wind up in the analysis bin for another library without specifying how.” *Id.* In other words, Dr. Adalsteinsson admitted that EX2005 is irrelevant to his assertion that “cross-over error for flow-cell based NGS was not thought to be a problem.” *See* EX2001, ¶114.

55. Respectfully, EX2005 does not even support his deposition assertion that “sequences from one library” do not “wind up in the analysis bin for another library.” EX1055 (Adalsteinsson Tr.), 25:22-26:15. In this article, genomic sequences bound by transcription factors were barcoded and sequenced using an Illumina sequencer. EX2005, Abstract. The reported data reveal that the level of signal “crossover” from one library to another would need to be exceptionally high to even be observable. For example, Figure 2 of EX2005 illustrates extensive noise in the background distribution (the scales of the y-axes are between 0 and 800 units) with substantial variation and inconsistency rather than a flat distribution.

Measuring how many “sequences from one library” did or did not “wind up in the analysis bin for another library” in this context would have been very difficult and is effectively unmeasured. Figure 3B, relied on by the authors to evaluate signal “crossover” from one library to another, illustrates this difficulty. *See* EX2005, 4. Even for the sample with the strongest signal, Cse4, the strongest signal measured about 600 units but there were multiple other signals reaching 100 units or above. Because of this high background, it would be nearly impossible to distinguish which signal and what fraction of that signal were derived from another library. The only conclusion the authors reached was that sequence reads from one sample do not appear in sequence reads from another sample in a way that affected the conclusion of the study; it is never measured directly.

56. It is also notable that Figure 3B was generated using sequences that were *mappable* to reference sequences in the genome. EX2005, 14 (“Signal tracks were produced from uniquely-mapped reads.”). Specifically, in generating Figure 3B, the authors of EX2005 “discarded” various erroneous sequence reads, including sequence reads “lacking a fully-intact barcode” (*id.*, 13) and many sequence reads (up to 80%) that did not map to any reference sequence in the genome. *See id.*, 15 (Table 10); *id.*, 13 (fractioning sequence reads based on “mappability”). A POSA would have understood that these discarding steps would have removed many sequencing artifacts, like those observed and discarded in *Campbell* and *Maher*.

57. Simply put, EX2005 does not support Dr. Adalsteinsson's assertion that crossover was not believed to be a problem for NGS platforms such as Illumina sequencers. Nor does it support his argument that the field lacked awareness of the chimera problem. *See supra* § V.A.1.a. Instead, it confirms that researchers took deliberate steps to remove sequencing artifacts including artificial chimeras before conducting their analyses, supporting my opinion that chimera formation was a widely-appreciated concern by December 23, 2010.³

58. EXS2003-2004: Relying on EX2003 and EX2004, Dr. Adalsteinsson asserts that "cross-over errors arising during amplification . . . were not even recognized as an issue at the time of the invention." EX2001, ¶¶49, 136, 150. Setting aside that both documents published after December 2010, they do not support Dr. Adalsteinsson's characterization of the art as of December 2010.

³ Even if one assumes for the sake of argument that Dr. Adalsteinsson's interpretation of EX2005 is correct (and I believe it is not), his conclusion that "cross-over error for flow-cell based NGS was not thought to be a problem" (EX2001, ¶114) is contradicted by numerous prior art studies such as EX1005 (Gloor), EX1061 (Maher), EX1066 (Varela), and EX1057 (Campbell), which report on the formation of chimeric sequences arising during Illumina sequencing workflow.

59. EX2003 states that “[i]ndex misassignment between multiplexed libraries is a *known* issue that has impacted NGS technologies *from the time sample multiplexing was developed.*” EX2003, 1 (emphasis added). As both *Parameswaran* and *Gloor* are prior art examples of “sample multiplexing,” this contradicts Dr. Adalsteinsson’s assertion that the problem of chimera formation was unknown before December 2010. As discussed above, EX2004, which discusses dual-indexing, cited four of the prior art references (EXS1017-20) that I cited in my previous declaration (EX1003, ¶67) and explained that “PCR can produce chimeras by recombining different template molecules (a process often referred to as ‘jumping PCR’).” EX2004, 2. Moreover, in an actual prior art publication, the authors of EX2004 expressly disclosed the problem of “cross-contamination of samples through jumping PCR” in “pools of indexed libraries.” EX1015 (Meyer 2010), 9.

60. EXS2003, 2006-2007: Dr. Adalsteinsson asserts that “Illumina itself provided no indication that it was aware of cross-over error, or that it was an issue on Illumina sequencers.” EX2001, ¶¶128-129. In support, Dr. Adalsteinsson alleges that two Illumina whitepapers from 2008 and 2014 (EXS2006, 2007) discussed single-indexing, whereas a later Illumina whitepaper from 2018 (EX2003) discussed unique dual-indexing. In my view, one cannot rely on white papers to extrapolate what a company knew or did not know more broadly. Regardless, a POSA in December 2010 would have understood from multiple published studies that cross-

over error was a problem on Illumina sequencers, and a POSA therefore would have been motivated to apply *Parameswaran*'s unique dual-indexing method as discussed below. *See, e.g.*, EX1005 (Gloor), EX1061 (Maher), EX1066 (Varela), EX1057 (Campbell).

61. In sum, numerous prior-art studies confirm that chimeric molecules, including those “created during cluster amplification” on Illumina sequencers, were a known problem as of December 2010. Dr. Adalsteinsson's arguments to the contrary do not accurately reflect the state of the art, and when considered as a whole even the references he cites support my opinion that this was a known problem by December 2010.

2. Dr. Adalsteinsson Fails to Read *Parameswaran* from the Perspective of a POSA in December 2010

62. Dr. Adalsteinsson asserts that a POSA would not have been motivated to use *Parameswaran*'s unique dual-indexing method to reduce cross-over error. EX2001, ¶¶116-127. But a POSA would have recognized *Parameswaran*'s method as a suitable solution to the known problem of cross-over error.

a. *Parameswaran* Provided a Solution to the Known Problem of Chimeric Molecules

63. As discussed above, PCR amplification was known to create erroneous, artificial combinations of template nucleic acid molecules that were not present in the starting samples. *See supra* § V.A.1. A POSA would have been motivated to

identify and discard sequences from chimeric molecules formed from different samples regardless of when they arise—i.e., before or after PCR amplification on the flow cell—because these chimeric artifacts do not represent the sequences as originally obtained from the biological samples. *See supra* § V.A.1; EX1003, ¶160.

64. In December 2010, software was often used to identify these artificial chimeras. *See* EX1003, ¶161. The process typically required sophisticated bioinformatic analyses. EX1063 (Smyth), 2 (There were “the plethora of bioinformatic tools that are available for [identifying and removing artificial chimeras]”). For example, EX1061 (Maher) details several computational filters applied to detect and remove artificial chimeras. Similarly, *Gloor* (EX1005) discloses a bioinformatic approach to identify chimeric sequences. However, these methods were not universally applicable. For example, when the sequences of interest were previously unknown, or when reference sequences were not available for alignment, such strategies were inadequate. *See, e.g.*, EX2010, 4-5 (Some “errors can be easily identified and eliminated when closely related sequences are available for comparison, but may represent a more serious problem when this is not the case.”). Further, if chimeras were formed between two highly similar templates, bioinformatic tools might not identify them as an artificial chimera. EX1063 (Smyth), 2 (a bioinformatics approach “is likely to overlook more subtle naturally derived chimeras that occur between highly similar sequences”). “Unfortunately,

highly similar sequences are the most likely to form chimeras during PCR, meaning the most commonly formed chimeras are the most difficult to computationally detect.” *Id.* This reality underscored the practical need for alternative approaches.

65. *Parameswaran* offered a solution to this problem. *See* EX1003, ¶161. Using a unique dual-indexing method, *Parameswaran* demonstrated the ability to identify chimeric sequences created during sequencing workflow from pooled samples based on mismatched barcode motifs—i.e., barcode motifs that should not appear together because they originated from different samples. *Id.*, ¶¶161, 126-131; *see* EX1004 (*Parameswaran*), 12 (identifying sequences with barcodes “from another sample set”), 7-8 (characterizing the occurrence of sequences with mismatched barcode motifs as “false-discover[ies]”). Notably, *Parameswaran*’s unique dual-indexing methodology was not limited to identifying chimeric sequences formed at a specific stage of the sequencing workflow. Instead, it could detect inter-sample chimeric sequences that occurred any time after barcoding, providing even more power to the technology. *See infra* § V.C. As such, using *Parameswaran*’s method, a POSA performing multiplex sequencing on an Illumina platform would have been able to identify and discard sequences from chimeric molecules made of templates from different samples introduced during a sequencing workflow, including those created during cluster amplification.

b. Dr. Adalsteinsson's Contrary Arguments Ignore the Perspective of a POSA

66. Dr. Adalsteinsson asserts that a POSA would not have appreciated the applicability of *Parameswaran's* unique dual-indexing method because the reference “does not mention ‘chimeric’ or ‘chimeric molecules’ once.” EX2001, ¶117. I disagree. As explained above, a POSA in December 2010 would have understood that chimeric molecules would emerge at multiple stages during the sequencing workflow, including—but not limited to—(1) during PCR amplification that occurs after sample pooling and (2) during bridge amplification on the Illumina flow cell. *See supra* § V.A.1. The absence of an express reference to amplification after sample pooling does not diminish the relevance or applicability of *Parameswaran's* method. Rather, a POSA would have recognized that the method's core principle—identification and elimination of sequences with mismatched barcode motifs—would apply to chimeric molecules, including those formed through cross-over error.⁴ By 2010, a POSA would have readily understood the

⁴ Notably, one of the causes identified in *Parameswaran* for sequences with mismatched barcode motifs was “cross-contamination of template DNAs in the second round of PCR.” EX1004 (*Parameswaran*), 7. A POSA would have understood that this is equivalent to “cross-contamination of template DNAs” during

value of *Parameswaran*'s method, especially when *Gloor*, the other reference in the prior-art combination, specifically raised concerns about chimera formation during NGS on the Illumina platform. EX1005 (*Gloor*), 8 ("Chimeric sequences can arise during PCR or because of overlapping spots on the solid support when imaged during DNA sequencing.").

67. Dr. Adalsteinsson also speculates that "the mismatched error-free barcode motifs observed in *Parameswaran* were not a result of cross-over." EX2001, ¶116. According to Dr. Adalsteinsson, this is because a POSA would have understood 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." *Id.* I again disagree.

68. As Dr. Adalsteinsson noted, "six of the forward barcodes are exact reverse complements of six reverse barcodes" in *Parameswaran*. EX2001, ¶116; EX1004 (*Parameswaran*), 7. This created sequence reads from different samples that appeared to contain an identical barcode, even though the initial barcodes used for those samples were different. See EX1004 (*Parameswaran*), Figure 1 (depicting

cluster amplification because the root cause leading to the formation of chimeric molecules between the two sources is identical, i.e., PCR-induced artifacts created between different template nucleic acids.

sequencing reads generated from the forward and reverse directions). Samples 454-66 and 454-69 are good examples. The 5'-end barcode assigned to Sample 454-66 is ACGTTACCGT, which is the exact reverse complement of the 3'-end barcode assigned to Sample 454-69, ACGGTAACGT. EX1004 (Parmeswaran), 16 (Suppl. Table 1B). In other words, if you were to make the complementary strand for those barcode sequences and read it in reverse, it would appear identical to another barcode.

69. As *Parmeswaran* observed, this configuration created ambiguous sample assignments because the 10-nt barcodes alone were insufficient to distinguish a correct pair of barcodes assigned to a single sample versus an incorrect pair of barcodes assigned to two different samples. See EX1003, ¶¶83, 127. To address this ambiguity, *Parmeswaran* combined the 10-nt barcodes with three bases from the adjoining linker (forming a “barcode motif”). EX1004 (Parmeswaran), 7; EX1003, ¶¶83, 127. This made each barcode motif unique to each sample, because neither the sequence nor its reverse complement was repeated across different samples. See EX1004 (Parmeswaran), 16 (Suppl. Table 1B). Importantly, while this configuration reduced the number of sequences with conflicting barcodes, it did **not** eliminate them.

70. Overall, the fact that the number of sequences with mismatched barcodes decreased as the barcodes became truly unique demonstrates the power of

Parameswaran's unique dual-indexing method to identify erroneous, chimeric sequences and ensure accurate assignment of the remaining sequences to correct samples. The fact that multiple sequences with mismatched barcode motifs still remained after removing sample ambiguity caused by reverse complementary barcodes indicates that sequences with mismatched barcode motifs would have included sequences from chimeric molecules created between templates from two different samples.

71. Dr. Adalsteinsson offers alternative theories regarding the origin of the chimeric sequences in *Parameswaran*; some are implausible at best while others are immaterial. For instance, Dr. Adalsteinsson theorizes 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.” EX2001, ¶121. But *Parameswaran* explains that its barcodes were designed to “differ[] in at least 4nt,” and moreover states that error-free barcodes were required for its analysis. EX1004 (*Parameswaran*), 4. Thus, it is extremely unlikely that “errors in barcode synthesis” were the source of sequences with mismatched barcode motifs. Further, contrary to Dr. Adalsteinsson’s argument, *Parameswaran* used barcode motifs that were unique from each other. *Id.*, 7.

72. In addition, if, as Dr. Adalsteinsson theorizes, mismatched barcode motifs were caused by errors in barcode synthesis or sequencing, a POSA would

have expected that adding bases from the adjoining cloning linkers would not reduce the number of sequences with mismatched barcodes. Adding additional bases from the adjoining cloning linkers made the barcode motifs truly unique, but it would not have corrected any barcode synthesis errors in the initial barcodes. If the source of the error were in synthesizing or sequencing the barcode sequence, there would be no way to eliminate that error by sequencing additional, adjoining bases as part of a “barcode motif.” Thus, a POSA would have understood that sequences with mismatched barcode motifs in *Parameswaran* were not likely a result of barcode synthesis or sequencing error.

73. Dr. Adalsteinsson also speculates that “mismatched error-free barcodes [in *Parameswaran*] 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.” EX2001, ¶¶122-123. It is unclear how Dr. Adalsteinsson could determine that this was “likely,” because sequence identifiers alone cannot distinguish between chimeric molecules arising from contamination events, during pooling, amplification for cluster generation, or sequencing steps. *See supra* §§ IV.A, V.A.1; EX1055 (Adalsteinsson Tr.), 94:21-97:10 (“That’s something [that] can be hard to prove.”). Moreover, as discussed below, there are many applications where sequencing and sample assignment accuracy is important. For those applications, regardless of how

sequences with mismatched barcodes were made—whether through “contamination events occurring upstream,” during “the second round of PCR,” or anywhere in between—a POSA would have been motivated to discard *all* of those sequences because they do not accurately represent template nucleic acids present in the original samples and cannot accurately be assigned.

74. Citing EX2009, Dr. Adalsteinsson argues that, because “the second round of PCR . . . creates increased risk for contamination,” a POSA would have understood that avoiding the second round of PCR would eliminate chimera formation during sequencing workflow. EX2001, ¶¶102-104. But EX2009 does not teach or suggest that one could avoid multiple rounds of PCR in any and all situations (in most cases, as in *Gloor*, multiple rounds of PCR were necessary). A POSA would have viewed eliminating the second round of PCR for any and all situations untenable, especially when many multiplex sequencing applications relied on scarce genetic materials. EX1063, 2 (Avoiding PCR is “often impractical or impossible as PCR is invaluable when only small quantities of DNA are available for analysis.”). Further, EX2009 sequenced dual-indexed libraries using pyrosequencing and evaluated sequences that appeared to include the expected barcodes on both ends. EX2009, 6. But the study also explains that only a fraction of the sequences were actually evaluated because the sequence read lengths were insufficient to detect both barcodes. *Id.* (“Indeed, on average only 13.6% of the reads evidence MID sequences

[barcodes] on both sides of the read”). As shown in *Parameswaran*, the occurrence rate of sequences with mismatched barcodes in pyrosequencing can be as low as three sequences among 237,340 sequences. EX1004 (*Parameswaran*), 7 (Table 3). The mere fact that a small fraction of sequence reads did not contain mismatched barcodes would not have indicated to a POSA that avoiding the second round of PCR would eliminate chimeric sequences.

75. Lastly, Dr. Adalsteinsson asserts that a POSA would not have used *Parameswaran*’s unique dual-indexing method because “the false assignment rate of NGS sequencers could be sufficiently addressed with a single error-free index.” EX2001, ¶¶125-126 (citing EXS2010-2012). To support his opinion, Dr. Adalsteinsson quotes from EX2010, stating that “errors can be easily identified and eliminated when closely related sequences are available for comparison.” *Id.* But as Dr. Adalsteinsson confirmed in his deposition, he did not accurately describe what EX2010 actually reported, because he left out the underlined portion of the sentence: “These errors can be easily identified and eliminated when closely related sequences are available for comparison, but may represent a more serious problem when this is not the case.” EX2010 (Meyer 2008), 4-5. As Dr. Adalsteinsson further confirmed in his deposition, the “errors” in this sentence are not even about “false-assignment rates,” but about “specific types of errors” such as single base-pair insertion and deletion errors. EX1055 (Adalsteinsson Tr.), 116:1-117:6. EX2010 further

acknowledges that there are more “serious” errors that single-indexing approach may not sufficiently address. EX2010 (Meyer 2008), 4-5. In any event, I disagree with Dr. Adalsteinsson’s view that a single-indexing approach and the associated false assignment rate was adequate for any and all situations.

76. At the relevant time, a POSA would have recognized the benefits and limitations of single-index and unique dual-index schemes in multiplex sequencing, and would have selected the appropriate method based on practical considerations such as cost, sequencing scale, and the required level of accuracy. For example, a POSA would have understood that sequencing analyses based on single-indexing, like those used in the references relied on by Dr. Adalsteinsson (e.g., EXS2010-2012), could be satisfactory for applications where a lower level of accuracy in sample tracking and error identification was acceptable. In contrast, a POSA would have understood that the unique dual-indexing approach used in *Parameswaran* provided a higher level of confidence in sample tracking and error identification, and would have been desirable in applications where greater accuracy was important.

77. Indeed, a POSA would have understood that sample tracking and sequence fidelity are critical in many situations. For example, in cancer research, low-frequency somatic mutations can have profound biological consequences. EX1061 (Maher), 1 (“the identification of novel gene fusions in a broad range of cancers is of enormous therapeutic significance”). In ancient DNA research, the

observation of a single sequence may be taken as evidence for DNA survival or presence of chimeric contamination. EX1063 (Smyth), 1 (Chimera formation is a “common problem for population genetic and molecular evolution analyses as artificial chimeras can alter the shape of phylogenetic trees” or “inflate measurements of diversity.”).

78. In the clinical setting, accurate and affordable molecular diagnostics were paramount, and researchers in December 2010 recognized that *Parameswaran*’s method could enable this. EX1056 (Anderson), 19 (high-throughput sequencing “for clinical diagnostics becomes significantly more challenging when one considers both the reagent cost and technical as well as computational expertise required.”), 18 (“Coupled with the use of barcoded amplicon primers to multiplex multiple patients in a single run [citing *Parameswaran*], this approach may become one of the first applications of high-throughput sequencing to be adopted by clinical molecular laboratories.”). In his deposition, Dr. Adalsteinsson also acknowledged that, as of December 2010, many applications such as cancer research required highly accurate sample assignments. EX1055 (Adalsteinsson Tr.), 113:16-114:10. Thus, I disagree with Dr. Adalsteinsson that a POSA would have believed that a single index would “sufficiently” address the false assignment rate of NGS sequencers.

3. Dr. Adalsteinsson’s Analysis of the “Discarding” Step Ignores Important Disclosures in *Parameswaran* and *Gloor*

79. Dr. Adalsteinsson contends that “neither *Parameswaran* nor *Gloor* disclose or render obvious ‘discarding sequence reads . . . that contain said combination of the identifier sequences that are different than any of said distinct pairs.’” EX2001, ¶148. As explained below, I disagree.

a. *Parameswaran* Renders the “Discarding” Step Obvious

80. Dr. Adalsteinsson does not dispute that *Parameswaran* identified sequences with mismatched barcode motifs. See EX2001, ¶¶140-144. Nevertheless, he argues that *Parameswaran* does not teach or suggest the “discarding” step because: (1) “a POSA would not see a need to identify and discard sequences containing a mismatched pair of barcode motifs” in view of *Parameswaran*’s “low-false discovery rate”; and (2) “*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.” *Id.* I disagree on both issues, both because he fails to consider the disclosure of *Parameswaran* from the perspective of a POSA and because he misreads the reference.

81. In December 2010, it was well understood that erroneous sequences such as chimeric sequences could be kept or discarded depending on the sequencing project. Thus, even assuming for the sake of argument that *Parameswaran* identified

chimeric sequences but did not discard them, a POSA would have known to discard them when data integrity and accurate sample attribution were important. *Compare* EX1004 (Parameswaran), 8 (*Parameswaran* validating sequences containing “both 5’ and 3’ barcode motifs” (i.e., sequences with a correct pair of identifiers) through placement “into the correct bin”) *with id.* (Parameswaran calling sequences with mismatched barcode motifs as “false-discover[ies],” constituting “misassignment[s]” of sequences to incorrect samples”).

82. Prior art studies such as *Campbell* discarded over 1.2 million chimeric sequences because the majority of those sequences were “presumably . . . induced by sequencing errors.” EX1057 (Campbell), 2. Prior art studies routinely discarded PCR-induced artifacts to improve sequencing accuracy, demonstrating that a POSA would have been motivated to discard erroneous chimeric sequences. EX1061 (Maher), 8 (artificial chimeras were “filter[ed]”); EX1066 (Varela), 10 (PCR-induced artifacts generated “within the bridge amplified cluster” “were excluded from analysis”); EX1057 (Campbell), 7 (PCR-induced artifacts generated “within the PCR colony” “were excluded from analysis”). This is consistent with *Parameswaran’s* statement that the purpose of the sequencing study can influence which sequences are discarded. EX1004 (Parameswaran), 8 (stating that, “[f]or our purposes, all sequences with a match to one of the two expected sequence motifs were considered as true members of the corresponding dataset.”).

83. Referring to this sentence in *Parameswaran* (“For our purposes . . .”), Dr. Adalsteinsson asserts that “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.” EX2001, ¶140 (citing EX1004 (Parameswaran), 8). But this sentence refers to “sequences with a match to one of the two **expected** sequence motifs,” not sequences with mismatched barcode motifs. EX1004 (Parameswaran), 8 (emphasis added). Sequences with mismatched barcode motifs, which *Parameswaran* calls “perfect, **unexpected** pairs,” differ from other errors because they provide conflicting information about sample assignment, and thus prohibit sample assignment. EX1004 (Parameswaran), 7. *Parameswaran*’s characterization of these “perfect, unexpected occurrences” as “false-discoveries” makes sense, because when a sequence has a 5’-end barcode motif that belongs to one sample and a 3’-end barcode motif that belongs to a different sample, there would have been no way of knowing which of the two conflicting barcode motifs provides correct sample information. Accordingly, this sequence could **not** be “assigned to 25 independent, uniquely barcoded libraries based on the presence of either a perfect forward or a perfect reverse barcode.” *Id.*, 1.

84. *Parameswaran*’s data analysis illustrates this point. EX1004 (Parameswaran), 20-31 (Suppl. Tables 4-5). Below, I reproduce *Parameswaran*’s

Supplemental Table 5D, which shows that sequences with perfect, *unexpected* motifs and sequences with imperfect, *expected* motifs are treated differently.

D. SEARCH MOTIF = Barcode (10 nt) + Cloning-Linker (3 nt)

	454-66	454-67	454-68	454-69	454-70	454-71	454-72	454-73	454-74	454-75	454-76	454-77	454-78	454-79	454-80	454-81	454-82	454-83	454-84	454-85	454-86	454-87	454-88	454-89	454-90	(close)	(no match)
454-66	13566	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	639	85
454-67	0	7459	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	221	33
454-68	0	0	13260	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	314	139
454-69	0	0	0	6518	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	108	115
454-70	0	0	0	0	9484	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	350	23
454-71	0	0	0	0	0	9397	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	297	55
454-72	0	0	0	0	0	0	12302	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	264	54
454-73	0	0	0	0	0	0	0	7273	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	145	32
454-74	0	0	0	0	0	0	0	0	13923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	211	51
454-75	0	0	0	0	0	0	0	0	1	15060	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	418	97
454-76	0	0	0	0	0	0	0	0	1	0	10266	0	0	0	0	0	0	0	0	0	0	0	0	0	0	308	50
454-77	0	0	0	0	0	0	0	0	0	0	0	9602	0	0	0	0	0	0	0	0	0	0	0	0	0	241	49
454-78	0	0	0	0	0	0	0	0	0	0	0	0	12082	0	0	0	0	0	0	0	0	0	0	0	0	226	34
454-79	0	0	0	0	0	0	0	0	0	0	0	0	0	10349	0	0	0	0	0	0	0	0	0	0	0	269	48
454-80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10019	0	0	0	0	0	0	0	0	0	0	294	153
454-81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6799	0	0	0	0	0	0	0	0	0	200	42
454-82	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	11674	0	0	0	0	0	0	0	0	357	28
454-83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12141	0	0	0	0	0	0	0	292	43
454-84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14049	0	0	0	0	0	0	405	77
454-85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7632	0	0	0	0	0	155	44
454-86	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8838	0	0	0	0	0	250	77
454-87	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	10431	0	0	0	0	390	57
454-88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10395	0	0	0	202	65
454-89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8148	1	201	56	
454-90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5028	136	27
(close)	369	167	210	122	237	154	317	172	268	280	209	215	299	319	340	221	325	268	346	218	198	269	258	191	126		
(no match)	37	26	76	48	28	20	48	29	64	46	42	36	78	59	60	47	58	35	52	26	77	27	70	59	29		

Figure 3. *Parameswaran's* Supplemental Table 5D (Original)

85. As *Parameswaran* explains, the rows denote barcode motifs at the 5'-end of the sequence reads, and the columns denote barcode motifs at the 3'-end of sequence reads. EX1004 (*Parameswaran*), 11. Each pair of barcode motifs uniquely corresponds to one particular sample (numbered Samples 454-66 through 454-90). *Id.*, 16-17 (Suppl. Table 1B). As summarized below, *Parameswaran* explains that the sequence reads fall into three categories. For ease of discussion, I annotated certain cells of Supplemental Table 5D as shown below:

D. SEARCH MOTIF = Barcode (10 nt) + Cloning-Linker (3 nt)

	454-66	454-67	454-68	454-69	454-70	454-71	454-72	454-73	454-74	454-75	454-76	454-77	454-78	454-79	454-80	454-81	454-82	454-83	454-84	454-85	454-86	454-87	454-88	454-89	454-90	(close)	(no match)
454-66	13566	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	639	85
454-67	0	7459	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	221	33
454-68	0	0	13260	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	314	139
454-69	0	0	0	6518	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	108	115
454-70	0	0	0	0	9484	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	350	23
454-71	0	0	0	0	0	9397	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	297	55
454-72	0	0	0	0	0	0	12302	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	264	54
454-73	0	0	0	0	0	0	0	7273	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	145	32
454-74	0	0	0	0	0	0	0	0	13923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	211	51
454-75	0	0	0	0	0	0	0	0	1	15060	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	412	97
454-76	0	0	0	0	0	0	0	0	1	0	10266	0	0	0	0	0	0	0	0	0	0	0	0	0	0	308	50
454-77	0	0	0	0	0	0	0	0	0	0	0	9602	0	0	0	0	0	0	0	0	0	0	0	0	0	241	49
454-78	0	0	0	0	0	0	0	0	0	0	0	0	12082	0	0	0	0	0	0	0	0	0	0	0	0	226	34
454-79	0	0	0	0	0	0	0	0	0	0	0	0	0	10349	0	0	0	0	0	0	0	0	0	0	0	269	48
454-80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10019	0	0	0	0	0	0	0	0	0	0	294	153
454-81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6799	0	0	0	0	0	0	0	0	0	200	42
454-82	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	11674	0	0	0	0	0	0	0	0	357	28
454-83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12141	0	0	0	0	0	0	0	292	43
454-84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14049	0	0	0	0	0	0	405	77
454-85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7632	0	0	0	0	0	155	44
454-86	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8838	0	0	0	0	250	77
454-87	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	10431	0	0	0	390	57
454-88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10395	0	0	202	65
454-89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8148	1	201	56
454-90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5028	136	27
(close)	369	167	210	122	237	154	317	172	268	280	209	215	299	319	240	221	325	268	346	218	198	269	258	191	126		
(no match)	37	26	76	48	28	20	48	29	64	46	42	36	78	59	60	47	58	35	52	26	77	27	70	59	29		

Figure 4. *Parameswaran's* Supplemental Table 5D (Annotated)

86. Perfect, Expected Barcode Motifs: Many sequences “contain[ed] identifiable, perfectly matched 5’ and 3’ search motifs.” EX1004 (*Parameswaran*), 11. These sequences can be unambiguously assigned to specific samples because both barcode motifs identify the same sample. For example, there were 6,518 sequences (purple) that contain a 5’-end barcode motif corresponding to Sample 454-69 (blue) and a 3’-end barcode motif also corresponding to Sample 454-69 (red). These 6,518 sequences were “unambiguous[ly]” assigned to Sample 454-69. *See id.*, 8. The sequences that fall into this category appear in the cells of the above table where the row for a given sample intersects the column for that same sample, forming a diagonal line extending down from the upper-left-most cell.

87. Imperfect, Expected Barcode Motifs: Some sequences had “a perfect 5’ motif, but have imperfect 3’ motifs.” EX1004 (Parameswaran), 11. If the imperfect 3’ motifs had high sequence identity to a 3’-end barcode motif that was expected to be found with the 5’-end barcode motif, they were placed in a column marked as “(close).” *Id.* Otherwise, they were placed in a column marked as “(no match).” *Id.* Similarly, some sequences had “a perfect 3’ motif, but have imperfect 5’ motifs.” *Id.* If the imperfect 5’ motifs had high sequence identity to a 5’-end barcode motif that was expected to be found with the 3’-end barcode motif, they were placed in a row marked as “(close).” *Id.* Otherwise, they were placed in a row marked as “(no match).” *Id.*

88. For the purposes of the study, these sequences could be reasonably assigned to specific samples because the barcode motifs did not provide conflicting sample information. For example, sequences highlighted in green contained a 5’-end barcode motif corresponding to Sample 454-69 and a 3’-end barcode motif that had varying “imperfect” correspondence to the other expected barcode, but did not provide conflicting information for a different sample. Similarly, sequences highlighted in yellow contained a 3’-end barcode motif corresponding to Sample 454-69 and a 5’-end barcode motif that had varying “imperfect” correspondence to the other expected barcode, but did not provide conflicting information for a different sample. As *Parameswaran* explains, for their purposes, because the two

barcode motifs did not conflict with each other, these sequences “could be assigned” to Sample 454-69 based on “a perfect forward or a perfect reverse barcode.” *See* EX1004 (Parameswaran), 1.

89. *Perfect, Unexpected Barcode Motifs*: The remaining sequences were “*false-assigned* sequences (i.e., number of sequences that are in *off-diagonal squares*, excluding the number of imperfect matches).” EX1004 (Parameswaran), 11 (emphasis added). These sequences contained mismatched barcode motifs, i.e., a 5’-end barcode motif that belonged to one sample and a 3’-end barcode motif that belonged to a different sample. For example, one sequence (grey arrow) contained a 3’-end barcode motif corresponding to Sample 454-6~~9~~6 and a 5’-end barcode motif corresponding to Sample 454-6~~6~~9. Three sequences (grey arrow) contained a 3’-end barcode motif corresponding to Sample 454-6~~9~~6 and a 5’-end barcode motif corresponding to Sample 454-6~~7~~9. As *Parameswaran* explains, these samples with mismatched 5’ and 3’ barcode motifs were “false-discover[ies].” *See id.*, 1, 8. When presented with a similar example, Dr. Adalsteinsson agreed that sequences with mismatched barcodes cannot be assigned to any sample. EX1055 (Adalsteinsson Tr.), 52:15-53:5 (“There wouldn’t be a sample to assign it to”). Notably, these sequences are *greyed out* in *Parameswaran*, indicating that they were discarded from the sequencing analysis because they could not be assigned to any one sample.

At a minimum, all of this indicates it would have been obvious to discard these sequences. *Parameswaran* renders the “discarding” step obvious.

b. *Gloor* in Combination with *Parameswaran* Renders the “Discarding” Step Obvious

90. Dr. Adalsteinsson also acknowledges that many chimeric sequences were removed from the sequencing analysis in *Gloor*. EX2001, ¶¶146-147. But Dr. Adalsteinsson contends that this does not render the “discarding” step obvious because “*Gloor* does not specify whether these incorrect chimeric sequences arose during amplification after sample pooling.” *Id.* He further argues that *Gloor*’s “non-unique dual indexing method” would result “in assigning sequence reads from a chimeric molecule to a sample instead of discarding them.” *Id.*

91. On the contrary, *Gloor* buttresses the POSA’s motivation to identify and discard chimeric sequences. *First*, *Gloor* expressly recognizes that artificial chimeric sequences (detected via a bioinformatics analysis) could “arise during PCR” (EX1005 (*Gloor*), 8)) and cites EX1017 (Lahr). A POSA reading *Gloor* would have understood that these sequences arose during the sequencing workflow, including during PCR amplification on the Illumina flow cell. *See supra* § V.A.1. *Second*, *Gloor* explains that “[t]he dataset was used without further regard to chimeric sequences,” because many such sequences had *already* been eliminated from the dataset. EX1005 (*Gloor*), 8. This was achieved through stringent filtering criteria, including a “strict requirement for proper sequence tag and primer

sequences on the left and right ends” and the imposition of a “requirement for long overlapping segments of a defined length.” *Id.*, 22. *Third*, even after the stringent filtering criteria that eliminated many chimeric sequences from the dataset, *Gloor* still observed chimeric sequences. *Id.*, 8, 22. A POSA interested in sample tracking and sequence fidelity would therefore have been motivated to identify and discard chimeric sequences by adopting *Parameswaran*’s method to an Illumina sequencer. The “discarding” step would have been obvious.

4. Dr. Adalsteinsson’s Reasons for *Not* Combining *Parameswaran* and *Gloor* Lack Scientific Merit

92. In my previous declaration, I identified multiple reasons why a POSA would have been motivated to combine *Parameswaran* and *Gloor*. See EX1003, ¶¶156-164. Dr. Adalsteinsson identifies purported reasons why a POSA *would not* have combined these references. EX2001, ¶¶91-97, 131. As discussed below, I disagree with Dr. Adalsteinsson’s opinions.

93. *First*, Dr. Adalsteinsson asserts that a POSA “would not have seen any utility in *Parameswaran*’s pyrosequencing-tailored methods on Illumina sequencers” because Illumina sequencers “enhanced sequencing capacity beyond what *Parameswaran*’s method was capable of sequencing.” EX2001, ¶¶92-93. Dr. Adalsteinsson fails to acknowledge the sequencer-agnostic benefits conferred by *Parameswaran*, regardless of the efficiency and cost-savings associated with Illumina sequencers. A POSA would have recognized that *Parameswaran* addressed

the problem of sample tracking in multiplex sequencing. EX1004 (Parameswaran), 1. As I explained previously, this issue was not unique to the Roche 454 platform—it was equally relevant to Illumina sequencing systems, regardless of distinctions in error rates or sequencing throughput between the platforms. *See* EX1003, ¶120. Moreover, the formation of chimeric molecules was recognized as an inherent risk in multiplex sequencing workflows, irrespective of the specific platform employed. *See supra* § V.A.1; EX1003, ¶¶67-69, 58-61, 23. Indeed, *Gloor* explicitly documents the occurrence of chimeras following multiplexed sequencing conducted on an Illumina instrument. *See* EX1005 (Gloor), 8, 22.

94. By December 2010, it was widely appreciated that the barcoding methodology of *Parameswaran* could be applied to other NGS sequencing platforms, including Illumina sequencers. For example, in a section entitled “Next-Generation Sequencing Strategies,” *Pfrender* (EX1062) discusses both Roche 454 and Illumina sequencers. *Id.*, 11. It cites *Parameswaran* to support the statement that “[a]pplying a tagging strategy to uniquely identify samples would allow the combination of multiple whole community assays in a single run, thereby reducing the per sample cost.” *Id.* This September 2010 reference thus directly connects NGS—expressly including Illumina sequencers—to the tagging strategy disclosed in *Parameswaran* and identifies increased throughput and decreased cost as advantages of the approach. *Id.*

95. *Anderson* (EX1056) also expressly acknowledged the broad applicability of *Parameswaran*'s method to NGS platforms. *Anderson* surveyed "next generation sequencing platforms," including the Illumina Genome Analyzer. *Id.*, 7-8. *Anderson* noted that "next-generation sequencing instruments offer a novel approach to detect infectious organisms and minimal residual diseases (MRD)." *Id.*, 17. *Anderson* emphasized that "MRD detection is important for many diseases" such as leukemia and lymphoma. *Id.*, 17-18. *Anderson* further noted that "[h]igh-throughput sequencing methods can improve MRD detection," and "[c]oupled with the use of barcoded amplicon primers to multiplex multiple patients in a single run [in *Parameswaran*], this approach may become one of the first applications of high-throughput sequencing to be adopted by clinical molecular laboratories." *Id.*, 18. Like *Pfrender*, *Anderson* expressly cited *Parameswaran* and connected its barcoding strategy to high-throughput sequencing platforms such as Illumina sequencers. *Id.*

96. In *Wu* (EX1068), the authors noted that "[t]he pooling of DNA samples from the parental lines for the target population, combined with the high-throughput sequencing of a small fraction of the genome results in the identification of a large number of SNPs genome wide." *Id.*, 7. *Wu* explained that the "drawbacks of pooling the DNA samples" can be "overcome by using a barcoding technique." *Id.* Specifically, *Wu* cites *Parameswaran* for the proposition that "[i]t is not a problem to sequence many samples in a Solexa [Illumina] instrument run if using a barcoding

technique. It is becoming a standard approach to increase the number of samples run on high-throughput instructions.” *Id.*; EX1010 (Kircher), 5 (“Solexa sequencing” is “now called the Illumina Genome Analyzer”). Like *Pfrender* and *Anderson*, *Wu* thus expressly identifies the barcoding technique in *Parameswaran* as advantageous when combined with Illumina sequencers.

97. An inventor of the ’852 patent also discussed the applicability of *Parameswaran*’s method to all “ensemble” based sequencing methods—including Illumina sequencers. EX1059 (*Fuller*), 1 (authors including George M Church). *Fuller*, which is a 2009 prior-art article, cited *Parameswaran* in stating that a “streamlined, robust sample preparation” method would be “an effective technology to judiciously add sequence tags to the fragments of several different samples,” such that DNA samples could be combined and processed together, with the resulting sequences separated based on their tags. *Id.*, 3. *Fuller* then explained that *Parameswaran*’s method would reduce “the inefficiencies of ensemble-based SBS [sequencing by synthesis] architectures.” *Id.* The article further explained that ensemble-based strategies included any system involving the sequencing of multiple identical copies of a DNA (*id.*, 1), and categorized Illumina sequencers as an ensemble method because it uses amplified fragments (*id.*, Table 1, Figure 2).

98. Thus, contrary to Dr. Adalsteinsson's unsupported opinion, numerous prior art references reported direct connections between *Parameswaran's* methods and Illumina sequencers.

99. *Second*, Dr. Adalsteinsson asserts that, because *Parameswaran's* barcodes are "tailored" for pyrosequencing, its approach should be limited to pyrosequencing. EX2001, ¶¶91-93. This is inconsistent with a POSA's understanding by December 2010. The fact that *Parameswaran* validated its unique dual-indexing strategy using a Roche 454 sequencer would not have precluded a POSA from implementing the same approach on an Illumina platform. Indeed, as explained previously, designing barcode sequences for different sequencing platforms was a routine task as of December 2010. EX1003, ¶¶166-170; EX1055 (Adalsteinsson Tr.), 104:13-105:11 ("So by and large, I'd say that was known in the art."). For example, cross-platform adaptation had already been demonstrated in prior-art studies such as *Maughan* (EX1038), which reported the successful application of unique dual-indexing on both Roche 454 and Illumina sequencers using the same set of barcodes. EX1003, ¶170. A POSA applying *Parameswaran's* unique dual-indexing method to an Illumina sequencer would have known how to design appropriate barcodes compatible with Illumina's system.

100. The '852 patent further confirms that barcode design was not a barrier. EX1003, ¶167. The '852 patent does not specify particular barcode sequences.

Instead, it relies on the knowledge of a POSA, stating that “[n]umerous techniques are known for detecting sequence and for identifying barcodes and some are exemplified below. However, the exact means for detecting and compiling sequence data does not affect the function of the invention.” EX1001 (’852 patent), 3:44-48, 8:65-9:3 (citing prior art approaches for “designing sets of barcode sequences”), 20:37-41 (“The invention is not intended to be limited to use of the sequencing method outline above, as essentially any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used.”).

101. I also disagree with Dr. Adalsteinsson’s related argument that *Gloor* “cautioned” against applying pyrosequencing approaches to Illumina sequencers. EX2001, ¶94. The statement Dr. Adalsteinsson relies on from *Gloor* discusses analytical considerations specific to microbiome studies arising from the differing read structures produced by the Roche 454 system (which generated unidirectional reads) and the Illumina platform (which produced bidirectional, paired-end reads). EX1005 (*Gloor*) 2. *Gloor* provided solutions on how to adjust for these differences in data analysis. *See id.*, 6-7; EX1003, ¶¶92-95. A POSA, familiar with these distinctions, would have understood *Gloor* as offering clear guidance for adapting *Parameswaran*’s strategy to Illumina sequencing.

102. *Third*, Dr. Adalsteinsson asserts that a POSA would not have adopted *Parameswaran*’s unique dual-indexing strategy because both *Gloor* and

Parameswaran discussed a combinatorial dual-indexing strategy. EX2001, ¶¶95-96. I disagree with Dr. Adalsteinsson's overgeneralized assessment.

103. Similar to the pros and cons of dual-indexing and single-indexing discussed above, a POSA would have understood the trade-offs between unique and non-unique dual-indexing methods and would have selected the appropriate method based on practical considerations such as cost constraints, sequencing scale, the known properties of the sequencers, and the required level of accuracy. *See supra* §§ V.A.1-3. A POSA would have understood that *Parameswaran*'s unique dual-indexing system conferred superior sequencing fidelity by effectively minimizing a known source of error affecting multiplex sequencing. EX1003, ¶¶160-161. And it would have done so in a manner that was precise and broadly applicable, unlike pure bioinformatic approaches. For applications like clinical diagnosis, where sample tracking and sequence fidelity are important, a POSA would have been motivated to combine *Parameswaran*'s unique dual-indexing method with the Illumina sequencers of *Gloor*, which had "an exceedingly low error rate" and were "an approximate order of magnitude cheaper than the per-read cost of pyrosequencing." EX1005 (*Gloor*), 2.

104. *Lastly*, Dr. Adalsteinsson does not dispute that *Parameswaran* expressly suggested applying its method to "other sequencing platforms," including Illumina sequencers. EX1004 (*Parameswaran*), 1, 2, 8; EX1003, ¶162. Instead, Dr.

Adalsteinsson asserts that a POSA would have disregarded this express suggestion, because “the field had moved beyond Parameswaran’s method and already enabled multiplexing in Illumina sequencers.” EX2001, ¶131.

105. Dr. Adalsteinsson’s opinion is unfounded and inconsistent with the contemporaneous literature. Roche 454 pyrosequencing, used in *Parameswaran*, was still regarded as one of the major NGS platforms in 2010. *See, e.g.*, EX1011 (Metzker), Table 1; EX1010 (Kircher), Table 1. The specification of the ’852 patent also acknowledges that Roche 454 pyrosequencing was a suitable NGS option in 2010. EX1001 (’852 patent), 20:37-46. A follow-on product invented by the same team, with nearly identical parameters except for the detection systems, the Ion Torrent, is mentioned in the ’200 Patent specification. *Id.* Likewise, researchers did *not* move beyond *Parameswaran*’s method. Rather, researchers repeatedly recommended *Parameswaran*’s method for a variety of applications performed on NGS platforms such as Illumina sequencers. *See* EX1062 (Pfrender); EX1055 (Anderson); EX1068 (Wu); EX1059 (Fuller). Likewise, the fact that “multiplexing in Illumina sequencers” was enabled in December 2010 did not make *Parameswaran*’s method any less relevant because a POSA would have recognized the sequencer-agnostic benefits offered by *Parameswaran*’s barcoding method. If anything, *Parameswaran*’s method became *more* relevant because the generation of chimeric sequences during multiplex sequencing on Illumina sequencers had been

empirically demonstrated and the throughput of instruments was continuing to expand exponentially, driving larger and larger amounts of multiplexing. EX1005 (Gloor), 8, 22.

106. Dr. Adalsteinsson cannot dispute that Illumina sequencers were one of the major NGS platforms available in December 2010. EX1003, ¶49. Numerous studies as of December 2010 compared sequencing results from Roche 454 and Illumina platforms. *See* EX1015 (Meyer 2010), 1 (adapting protocols developed for Roche 454 to Illumina); EX1038 (Maughan), 1 (performing unique dual-indexed, multiplex sequencing using Roche 454 and Illumina sequencers); EX1058 (Delseny), 8 (“Many projects initiated with the 454 technology are being continued with Illumina or SOLiD platforms and will certainly use the latest methods for finishing.”). It would have been obvious for a POSA to follow *Parameswaran’s* express suggestion to apply its method to Illumina sequencers.

5. Dr. Adalsteinsson’s Focus on Absolute Number of Barcodes Ignores Context

107. Dr. Adalsteinsson criticizes me for “only focus[ing] on the cost of the sequencers themselves and ignore[ing] extra costs associated with *Parameswaran’s* unique dual indexes.” EX2001, ¶¶98-101; *see also id.*, ¶96. According to Dr. Adalsteinsson, “if the 272 samples that Gloor were sequenced with [sic] 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 . . . that Gloor actually uses.” *Id.*, ¶100.

108. I disagree with Dr. Adalsteinsson’s unnecessarily mechanical combination of *Parameswaran* and *Gloor*. Dr. Adalsteinsson appears to imply that “16 times more” barcodes required for sequencing “the 272 samples” in *Gloor* would have somehow dissuaded a POSA from using *Parameswaran*’s method. But nothing in the claims, the specification, or the prior art requires sequencing “272 samples.” As an illustration, consider a situation where 8 samples are sequenced. *Gloor*’s combinatorial dual-indexing method requires 6 barcodes (3 forward and 3 reverse barcodes). *Parameswaran*’s unique dual-indexing method requires 16 barcodes (8 forward and 8 reverse barcodes). Making and using 16 barcodes was certainly within the capacity of a POSA as of December 2010. *See* EX1005 (*Gloor*), 6 (Table 2). *Parameswaran* itself, a 2007 reference, made and used 64 barcode sequences. EX1004 (*Parameswaran*), 4. For a POSA in December 2010 working in the field of genomic sequencing, the standard scale for sequencing reactions was 12, 96, and 384, which corresponds to the standard microtiter plates routinely used in the field. Making or purchasing barcode-containing primers for this standard scale was done routinely. Moreover, once the primers were obtained, they could be used for thousands of PCR reactions, incurring a one-time cost in most cases.

109. Even if barcode synthesis cost or error was a relevant concern for some applications, a POSA would have understood that sample tracking and sequence fidelity are critical in many situations, including cancer research, population genetics, and clinical diagnosis. *E.g.*, EX1061 (Maher), 1 (“the identification of novel gene fusions in a broad range of cancers is of enormous therapeutic significance”); EX1063 (Smyth), 1 (Chimera formation is a “common problem for population genetic and molecular evolution analyses as artificial chimeras can alter the shape of phylogenetic trees” or “inflate measurements of diversity.”); EX1056 (Anderson), 19 (high-throughput sequencing “for clinical diagnostics becomes significantly more challenging when one considers both the reagent cost and technical as well as computational expertise required”); *see supra* §§ V.A.1-4. Reflecting this, researchers (including an inventor of the ’852 patent) recommended *Parameswaran’s* method for a variety of applications performed on NGS platforms such as Illumina sequencers. *See* EX1062 (Pfrender); EX1056 (Anderson); EX1068 (Wu); EX1059 (Fuller); *see supra* § V.A.4. Dr. Adalsteinsson disregards these factors, which would have strongly motivated a POSA to apply *Parameswaran’s* unique dual-indexing method to the Illumina sequencer in *Gloor*.

110. Finally, a POSA would have known how to choose appropriate barcode configurations depending on the goal of sequencing. Non-unique dual-indexing can still provide sample assignment advantages when the available barcode space is far

from saturated. For example, instead of doing 272 pairs, one could eliminate approximately 90% of chimeras made of templates from different samples using 50 forward and 50 reverse pairs, allowing 2,500 combinations for just 272 samples. In comparison, a POSA would have been able to identify and eliminate *all* chimeras made of templates from different samples using unique dual-indexing using the same number of barcodes but with reduced sample capacity. In my opinion, both were obvious design choices, not a dichotomic decision that a POSA needed to make based on the single parameter of barcode numbers and expense, as Dr. Adalsteinsson asserts.

6. Dr. Adalsteinsson's Argument on Reasonable Expectation of Success Is Flawed

111. Dr. Adalsteinsson reiterates that “the claims of the ’852 Patent require ‘a method for reducing template cross-over error’ and actions taken ‘to reduce cross-over error.’” EX2001, ¶132. He then contends that “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.” *Id.*, ¶¶133-134. Like his motivation arguments, Dr. Adalsteinsson's expectation of success arguments are premised on his incorrect characterization of the state of the art.

112. As I discuss above, by December 2010, it was known that chimeric sequences could form at various stages during the sequencing workflow, including but not limited to during cluster amplification on Illumina sequencers. *See supra* §§ IV.A, V.A.1. By then, a POSA also would have known that this occurred in the context of multiplex sequencing because templates from multiple samples were pooled and then PCR-amplified, sometimes using multiple PCR steps. *Id.* Even beyond the express recognition of artificial chimeric molecule formation set forth in multiple prior art references, *see supra* § V.A.1, there was no scientific basis for asserting that chimeric molecules would not form during cluster amplification on the flow cells of Illumina sequencers (as opposed to forming during any PCR amplification steps implemented before cluster formation) or among templates from different samples (as opposed to a single sample). *Id.*

113. Dr. Adalsteinsson also argues that a POSA would not have reasonably expected that *Parameswaran's* method would reduce cross-over error because *Parameswaran* “does not discuss the existence of cross-over at all.” EX2001, ¶¶135-136. As discussed above, the utility of *Parameswaran's* unique dual-indexing technique in detecting chimera artifacts would have been recognized by a POSA, regardless of whether or not the term “cross-over” is explicitly used; indeed, the presence of “perfect unexpected matches” would have been understood to include chimeric molecules resulting from PCR “cross-over” events. *See supra* §§ V.A.2-3.

The scientific literature and practice in the field made it clear that amplification-associated artifacts could occur at various stages, and *Parameswaran's* method designed to track barcodes and detect anomalous combinations of barcodes would be applicable for identifying cross-over across those stages. *Id.* Thus, a POSA would have had a reasonable expectation of success in applying *Parameswaran's* method, discarding anomalous combinations of sequence identifiers, and thereby reducing cross-over error. Tellingly, Dr. Adalsteinsson does not appear to dispute that *Parameswaran's* unique dual-indexing method would reduce cross-over error when applied on an Illumina sequencer. *See* EX2001, ¶¶135-136. Indeed, the method steps in *Parameswaran* are identical to those recited in the claims of the '852 patent, except for *Parameswaran's* use of beads versus Illumina flow cells.

114. Dr. Adalsteinsson also asserts that a POSA would not have expected that “using *Parameswaran's* UDIs would provide any benefit over *Gloor's* CDI method.” EX2001, ¶¶137-138. As I explained above, a POSA would have understood the trade-offs between unique and non-unique indexing schemes and would have selected the appropriate method based on practical considerations such as the required level of accuracy, sequencing scale, and cost constraints. *See supra* §§ V.A.1-5. Indeed, when it comes to accurate sample assignment, a POSA would have recognized that *Parameswaran's* unique dual-indexing method offers substantial benefits over *Gloor's* combinatorial dual-indexing method, as it would

allow for accurately identifying and discarding chimeric sequences that could lead to critical sample misassignment. *Id.*

115. In sum, it is my opinion that a POSA would have been motivated to combine the teachings of *Parameswaran* and *Gloor* and would have had a reasonable expectation of success in doing so to practice the methods recited in the challenged claims of the '852 patent. To a POSA in December 2010, claims 1-8 of the '852 patent would have been obvious over *Parameswaran* and *Gloor*.

B. Ground 2: Claims 1-8 Would Have Been Obvious Over *Parameswaran* and *Bentley*

116. Dr. Adalsteinsson asserts that, for “the same reasons [he] discuss[ed]” regarding the combination of *Parameswaran* and *Gloor*, a POSA would not have been motivated or had a reasonable expectation of success to combine *Parameswaran* and *Bentley*. EX2001, ¶¶149, 151. As discussed above and in my previous declaration, I disagree. *See supra* § V.A; EX1003, ¶¶207-212. A POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* with *Bentley*, just as a POSA would have been motivated, with a reasonable expectation of success, to combine *Parameswaran* and *Gloor*.

117. Dr. Adalsteinsson alleges that I failed to identify the claimed “discarding” step in *Parameswaran* or *Bentley*. EX2001, ¶152. As discussed above and in my previous declaration, *Parameswaran* identified and characterized sequences with mismatched barcode motifs as “false-discover[ies],” and greyed

them out from sample assignment. *See supra* § V.A.3; EX1003, ¶206. *Parameswaran* also clarified that sequences could be kept or discarded depending on the purpose of the sequencing reaction. EX1004 (*Parameswaran*), 8. Both options would have been obvious to a POSA in December 2010. *See supra* § V.A.3.

118. Dr. Adalsteinsson also attempts to dismiss *Bentley*, arguing that “*Bentley* does not have any discussion of multiplexing.” EX2001, ¶¶149-150. Dr. Adalsteinsson fails to acknowledge that, prior to December 2010, Illumina sequencers, as used in *Bentley*, offered greater speed, greater efficiency, and lower cost compared to Roche 454 sequencers, and that Illumina sequencers had been used for many purposes, including multiplex sequencing. *See, e.g.*, EX1005 (*Gloor*), EX1015 (*Meyer 2010*), EX1038 (*Maughan*); EX1055 (*Adalsteinsson Tr.*), 73:4-7 (“Q. So you agree that by 2010 it was known that an Illumina sequencer could be used for multiplex sequencing; correct? A. Yes, I agree.”). A POSA would have understood that the unique dual-indexing method of *Parameswaran* would improve the already-superior accuracy of Illumina sequencers because erroneous chimeric sequences can be readily identified and discarded. *See supra* §§ V.A.4-5; EX1003, ¶207. Combined with *Parameswaran*’s express suggestion for applying its method to other sequencing platforms (EX1004, 8), a POSA had multiple reasons to combine *Parameswaran* and *Bentley* and would have had a reasonable expectation of success

in doing so. Indeed, switching between sequencers for various applications was routine.

C. Claims 1-8 Would Have Been Obvious Under Either My Construction or Dr. Adalsteinsson's Construction

119. I disagree with Dr. Adalsteinsson's construction that the claimed "template cross-over error" or "cross-over error" is limited to sequences of chimeric molecules *created* during cluster amplification. *See supra* § IV. But as explained above, regardless of which construction is adopted, claims 1-8 of the '852 patent would have been obvious to a POSA as of December 2010.

120. My construction of the term "template cross-over error" encompasses sequences of chimeric molecules made from template nucleic acids derived from different samples, regardless of when and how these sequences arise during the sequencing workflow. This is consistent with the understanding of a POSA in December 2010 that erroneous, chimeric sequences could arise at various stages during multiplex sequencing, including from PCR amplification occurring prior to and on the flow cell. *Supra* §§ IV.A, V.A.1; *see* EX1055 (Adalsteinsson Tr.), 94:21-97:10.

121. As Dr. Adalsteinsson admitted, the specific cause of a chimeric sequence created during the sequencing workflow was "hard to prove." EX1055 (Adalsteinsson Tr.), 94:21-97:10. But regardless of when and how these chimeras were created, a POSA would have understood that they do not correspond to nucleic

acid templates in the original samples. As of December 2010, there were numerous applications that demanded highly accurate genotyping. *See supra* §§ V.A.1-5; EX1055 (Adalsteinsson Tr.), 113:16-114:10. For those applications, a POSA would have been motivated to discard these erroneous, chimeric sequences, which could compromise sample assignment and sequencing accuracy.

122. As discussed in my previous declaration and above, *Parameswaran's* unique dual-indexing method offered a solution to this problem of chimeric sequences between samples. *See supra* §§ V.A.2-5; EX1003, ¶¶137-139. Using this indexing strategy, *Parameswaran* was able to identify sequences with a pair of identifier sequences that were different than any of the assigned pairs, i.e., mismatched barcode motifs. *See id.*, ¶¶126-130. As discussed, a POSA would have understood that these sequences included inter-sample chimeras that could have formed at various stages during the sequencing workflow. *See id.*, ¶¶130, 148; EX1055 (Adalsteinsson Tr.), 94:21-97:10. Using mismatched barcode motifs as a proxy, a POSA would have been able to identify and discard the sequences of these chimeric molecules, regardless of when and how those sequences were formed during the sequencing workflow. Indeed, as Dr. Adalsteinsson testified, unique dual-indexing is “a powerful approach that could address other weaknesses in the sequencing work flow,” not limited to addressing sequences of chimeric molecules

created at a particular stage of sequencing workflow. EX1055 (Adalsteinsson Tr.), 98:3-14. Indeed, it was *Parameswaran* that showed this ability.

123. In my opinion, (1) a POSA would have been motivated to discard *all* sequences from chimeric molecules made of templates derived from different samples; and (2) *Parameswaran*'s method applied to *Gloor*'s or *Bentley*'s Illumina sequencers would have allowed a POSA to identify and discard *all* such sequences that arose from *any* stage during Illumina's sequencing workflow. Thus, under my construction, the claims of the '852 patent would have been obvious in December 2010.

124. Dr. Adalsteinsson's construction, which limits the term "template cross-over error" to sequences of chimeric molecules made from template nucleic acids derived from different samples created during cluster amplification, does not change that the claims of the '852 patent would have been obvious. As discussed above, as with any PCR-based amplification, the problem of chimeric sequences created during cluster amplification in Illumina sequencing was known in the art. *See supra* § V.A.1. As with PCR-based amplifications performed prior to bridge amplification, a POSA would have understood that these inter-sample chimeric sequences generated during cluster amplification do not correspond to template nucleic acids contained in the original samples. For applications that demanded highly accurate genotyping, a POSA would have been motivated to discard chimeric sequences

created during cluster amplification, just like other aberrant chimeric sequences arising at other stages of the sequencing workflow. A POSA would have been motivated, with a reasonable expectation of success, to address these chimeric sequences using *Parameswaran*'s unique dual-indexing method because of its ability to address sequences of chimeric molecules from different samples arising at *any* stage of the sequencing workflow, including cluster amplification.

125. The schematic drawing below illustrates that a POSA would have arrived at the claimed methods of the '852 patent under either construction. Regardless of whether a POSA was motivated to identify and discard chimeric sequences arising before or during cluster amplification, application of *Parameswaran*'s unique dual-indexing to the Illumina sequencer of *Gloor* or *Bentley* would have reduced cross-over error arising before and during cluster amplification.

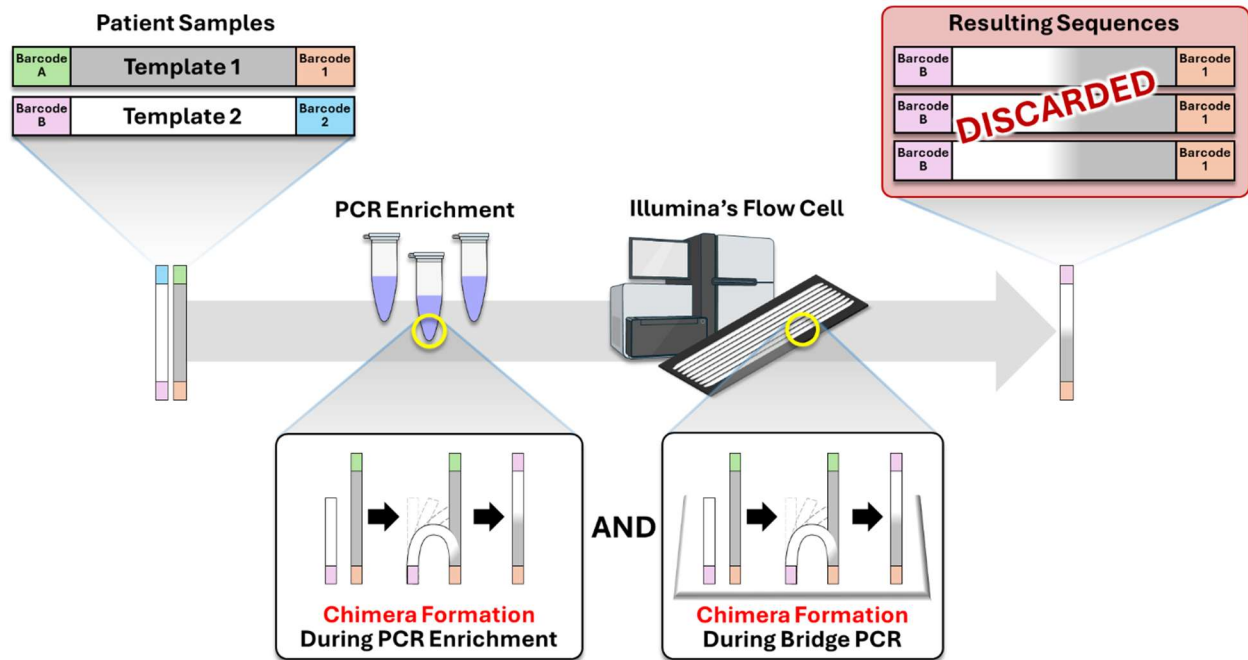


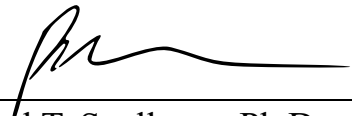
Figure 5. *Parameswaran's* Method Reduces Cross-Over Error Formed at Various Stages During Sequencing Workflow

VI. CONCLUSION

126. For the reasons discussed in my previous declaration and above, it is my opinion that claims 1-8 of the '852 patent would have been obvious to a POSA as of December 23, 2010.

127. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further, that these statements were made with knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001. I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Dated: July 3, 2025

By: 
Paul T. Spellman, Ph.D.

Appendix A

LIST OF MATERIALS CONSIDERED

Exhibit	Description
1001	U.S. Patent No. 11,041,852.
1003	Declaration of Paul T. Spellman, Ph.D.
1004	P. Parameswaran, et al., <i>A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing</i> , 35 NUCLEIC ACIDS RSCH. 19(e130) (2007).
1005	G.B. Gloor, et al., <i>Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products</i> , Cornell University Library arXiv:1007.5075v1 (2010).
1006	D.R. Bentley, et al., <i>Accurate whole human genome sequencing using reversible terminator chemistry</i> , 456 NATURE 6, 53-59 (2008).
1007	File History of U.S. Patent No. 11,041,852.
1010	M. Kircher & J. Kelso, <i>High-throughput DNA sequencing – concepts and limitations</i> , 32 BIOESSAYS 6, 524-536 (2010).
1011	M.L. Metzker, <i>Sequencing technologies—the next generation</i> , 11 NATURE REVIEWS GENETICS 31-46 (2010).
1012	J. Shendure & H. Ji, <i>Next-generation DNA sequencing</i> , 26 NATURE BIOTECH. 10, 1135-1145 (2008).
1015	M. Meyer & M. Kircher, <i>Illumina sequencing library preparation for highly multiplexed target capture and sequencing</i> , 2010, COLD SPRING HARBOR PROTOCOL 6 (2010).
1017	S.J.G. Lahr & L.A. Katz, <i>Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase</i> , 47 BIOTECHNIQUES 4, 857-866 (2009).
1018	A. Meyerhans, et al., <i>DNA recombination during PCR</i> , 18 NUCLEIC ACIDS RSCH. 7, 1687-1691 (1990).
1019	S. Paabo, et al., <i>DNA damage promotes jumping between templates during enzymatic amplification</i> , 265 J. OF BIOL. CHEM. 8, 4718-4721 (1990).
1020	S.J. Odelberg, et al., <i>Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I</i> , 23 NUCLEIC ACIDS RSCH. 11, 2049-2057 (1995).

Exhibit	Description
1021	D.W. Craig, et al., <i>Identification of genetic variants using bar-coded multiplexed sequencing</i> , 5 NATURE METHODS 10, 887-893 (2008).
1033	M. Dai, et al., <i>NGSQC: cross-platform quality analysis pipeline for deep sequencing data</i> , 11 BMC GENOMICS 11(Suppl 4):S7 (2010).
1034	S. Creer, et al., <i>Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises</i> , 19 MOLECULAR ECOLOGY (Suppl. 1) 4-20 (2010).
1038	P. Maughan, et al., <i>Single-Nucleotide Polymorphism Genotyping in Mapping Populations via Genomic Reduction and Next-Generation Sequencing: Proof of Concept</i> , 3 PLANT GENOME (3) 166-178 (2010).
1055	Deposition Transcript of Viktor Adalsteinsson, Ph.D.
1056	M.W. Anderson et al., <i>Next Generation DNA Sequencing and the Future of Genomic Medicine</i> , 1 GENES 38-69 (2010).
1057	P.J. Campbell et al., <i>Identification of Somatic Acquired Rearrangements in Cancer Using Genome-Wide Massively Parallel Paired-End Sequencing</i> , 40 NATURE GENETICS (6) 722-729 (2008).
1058	M. Delseny et al., <i>High Throughput DNA Sequencing: The new Sequencing Revolution</i> , 179 PLANT SCI. 407-422 (2010).
1059	C.W. Fuller et al., <i>The Challenges of Sequencing by Synthesis</i> , 27 NATURE BIOTECHNOLOGY (11) 1013-1023 (2009).
1060	I. Gorzer et al., <i>The Impact of PCR-Generated Recombination on Diversity Estimation of Mixed Viral Populations by Deep Sequencing</i> , 169 J. VIROLOGICAL METHODS 248-252 (2010).
1061	C.A. Maher et al., <i>Chimeric Transcript Discovery by Paired-End Transcriptome Sequencing</i> , 106 PNAS (30) 12353-12358 (2009).
1062	M.E. Pfrender et al., <i>Assessing Macroinvertebrate Biodiversity in Freshwater Ecosystems: Advances and Challenges in DNA-Based Approaches</i> , 85 Q. REV. BIOLOGY (3) 319-340 (2010).
1063	R.P. Smyth et al. <i>Reducing Chimera Formation During PCR Amplification to Ensure Accurate Genotyping</i> , 469 GENE 45-51 (2010).
1065	K.H. Taylor et al., <i>Next Generation Sequencing: Advances in Characterizing the Methyloome</i> , 1 GENES 143-165 (2010).

Exhibit	Description
1066	I. Varela et al., <i>Somatic Structural Rearrangements in Genetically Engineered Mouse Mammary Tumors</i> , 11 GENOME BIOLOGY R100, 1-12 (2010).
1067	N. Whiteford et al., <i>Swift: Primary Data Analysis for the Illumina Solexa Sequencing Platform</i> , 25 BIOINFORMATICS (17) 2194-2199 (2009).
1068	X. Wu et al., <i>SNP Discovery by High-Throughput Sequencing in Soybean</i> , 11 BMC GENOMICS 469, 1-12 (2010).
2001	Declaration of Viktor Adalsteinsson, Ph.D.
2003	Illumina White Paper – Effects of Index Misassignment on Multiplexing and Downstream Analysis, Illumina (2018), available at https://www.illumina.com/content/dam/illumina/marketing/documents/products/whitepapers/index-hopping-whitepaper-770-2017-004.pdf
2004	Kircher et al., <i>Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform</i> , Nucleic Acids Research 40(1):e3 (2011).
2005	Lefrancois et al., <i>Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing</i> , BMC Genomics 10:37 (2009).
2006	Illumina Application Note – High-Speed Multiplexed 16S Microbial Sequencing on the MiSeq® System, Illumina (2011), available at https://www.illumina.com/documents/products/appnotes/appnote_miseq_16S.pdf
2007	Illumina Sequencing – Multiplexed Sequencing with the Illumina Genome Analyzer System (2008), available at https://www.illumina.com/documents/products/datasheets/datasheet_sequencing_multiplex.pdf
2009	Maughan et al., <i>SNP Discovery via Genomic Reduction, Barcoding, and 454-Pyrosequencing in Amaranth</i> , The Plant Genome 2:260-270 (2009).
2010	Meyer et al., <i>Targeted high-throughput sequencing of tagged nucleic acid samples</i> , Nucleic Acids Research 35(15): e97 (2007).
2011	Wang et al., <i>A Low-Cost Library Construction Protocol and Data Analysis Pipeline for Illumina-Based Strand-Specific Multiplex RNA-Seq</i> , PLoS One 6(1): e26426 (2011)

Exhibit	Description
2012	Kenny et al., <i>Multiplex Target Enrichment Using DNA Indexing for Ultra-High Throughput SNP Detection</i> , DNA Research 18:31-38 (2010).