

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ILLUMINA INC.,
Petitioner,

v.

MOLECULAR LOOP BIOSCIENCES, INC.,
Patent Owner.

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

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
Claim 1[a]: “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest”	3
V. GROUND 1: CLAIMS 1-6 OF THE '200 PATENT ARE ANTICIPATED BY <i>GLOOR</i>	9
A. Dr. Adalsteinsson’s Opinion on Claim 1 Is Based Entirely on an Incorrect Claim Construction	9
B. Dr. Adalsteinsson’s Opinion on Claim 5 Is Contradicted by <i>Gloor</i> ’s Barcode Sequence Design	11
VI. GROUND 2: CLAIMS 1-6 OF THE '200 PATENT ARE ANTICIPATED BY <i>PARAMESWARAN</i>	13
A. Dr. Adalsteinsson’s Opinion on Claim 1 Ignores Important Disclosures of <i>Parameswaran</i>	13
B. Dr. Adalsteinsson’s Opinion on Claim 5 Ignores <i>Parameswaran</i> ’s General Teaching Regarding Barcode Design	20
C. Dr. Adalsteinsson’s Opinion on Claim 6 Adds an “Attachment” Limitation Not Required by the Claim	22
VII. GROUND 3: CLAIMS 1-6 OF THE '200 PATENT WOULD HAVE BEEN OBVIOUS OVER <i>PARAMESWARAN</i> AND <i>GLOOR</i>	24
A. Dr. Adalsteinsson’s Obviousness Opinion Focuses on Unclaimed Features	24

B.	Claims 1-4 and 6 Would Have Been Obvious over <i>Parameswaran</i> and <i>Gloor</i>	26
1.	Dr. Adalsteinsson’s Characterization of the State of the Art Is Inaccurate and Incomplete	26
2.	Dr. Adalsteinsson Fails to Read <i>Parameswaran</i> from the Perspective of a POSA in December 2010	43
3.	Dr. Adalsteinsson’s Reasons for <i>Not</i> Combining <i>Parameswaran</i> and <i>Gloor</i> Lack Scientific Merit	58
4.	Dr. Adalsteinsson’s Focus on Absolute Number of Barcodes Ignores Context	66
5.	Dr. Adalsteinsson’s Argument on Reasonable Expectation of Success Is Flawed	69
C.	Claim 5 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Gloor</i>	71
VIII.	GROUND 4: CLAIMS 1-6 OF THE ’200 PATENT WOULD HAVE BEEN OBVIOUS OVER <i>PARAMESWARAN</i> IN VIEW OF <i>BINLADEN</i>	72
A.	Claims 1-4 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Binladen</i>	72
1.	Dr. Adalsteinsson’s Opinion on the “Excluding” Step Is Incorrect	72
2.	Dr. Adalsteinsson’s Opinion Regarding Motivation to Combine <i>Parameswaran</i> and <i>Binladen</i> Reads These References in Isolation	75
B.	Claims 5 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Binladen</i>	77
C.	Claim 6 Would Have Been Obvious Over <i>Parameswaran</i> and <i>Binladen</i>	79

IX.	GROUND 5: CLAIM 6 OF THE '200 PATENT WOULD HAVE BEEN OBVIOUS OVER <i>PARAMESWARAN</i> , <i>BINLADEN</i> , AND <i>BENTLEY</i>	79
X.	CONCLUSION.....	81

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. 11,768,200 (“the ’200 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 ’200 patent should be evaluated from the perspective of a POSA as of December 23, 2010. *See* EX2001, ¶¶20-23, 43.

IV. CLAIM CONSTRUCTION

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

7. In opining on the anticipation and obviousness of the claims of the '200 patent, I read all claim terms as having their ordinary meaning as understood by a POSA as of December 23, 2010. EX1003, ¶133. Other than the term “homopolymer region” in dependent claim 5, Dr. Adalsteinsson states that, for all claim terms, he also applied their ordinary meaning as understood by a POSA. EX2001, ¶24. Likewise, in the section titled “CLAIM CONSTRUCTION,” Dr. Adalsteinsson did not propose construing any of the terms in claim 1. *Id.*

8. Dr. Adalsteinsson, however, contends that *Gloor* does not anticipate claim 1 of the '200 patent because, in his opinion, “a POSA would understand” the term “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest” to require “*each* identifier to be uniquely associated with the nucleic acid analyte of interest.” EX2001, ¶¶83-84 (emphasis added). Based on this implicit claim construction, Dr. Adalsteinsson concludes that “Gloor does not disclose the ‘detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest’ limitation” because *Gloor* used a combinatorial tagging method where each

pair of identifier sequences was unique to each sample but individual identifier sequences were used for multiple samples, *Id.*, ¶85.

9. I disagree with Dr. Adalsteinsson's interpretation of claim 1. The limitation at issue recites "detecting the presence of ***two or more identifier*** sequences that ***are uniquely associated*** with the nucleic acid analyte of interest." EX1001 ('200 patent), cl. 1 (emphasis added). This limitation requires only that the pair of "two or more identifier sequences" must be "uniquely associated" with the nucleic acid of interest—it does not recite that ***each*** identifier is uniquely associated with the nucleic acid analyte of interest, as Dr. Adalsteinsson suggests. A POSA reading this limitation would have understood that identifier sequences can be used for multiple nucleic acid analytes of interest provided that "two or more identifier sequences . . . are uniquely associated with that nucleic acid analyte of interest." This is consistent with the '200 patent specification's express statement that, "[w]hile the ***pair of barcodes associated with any given template is unique, individual barcode sequences can be used with multiple templates*** as long as no pair is used across two different samples." *Id.*, 2:50-53 (emphasis added). The specification further states that the "first and second identifiers may be the same or different, ***so long as the combination is unique*** with respect to the sample or sub-sample that one seeks to distinguish and/or validate." *Id.*, 3:20-23 (emphasis added).

10. Despite the clear language in both claim 1 and the specification, Dr. Adalsteinsson contends that: (1) “[t]he ’200 patent claims excluding sequences with an incorrect ‘pair of identifiers’ in its validating limitation”; and (2) a “POSA would understand that in order to validate a sequence by excluding sequences with ‘an incorrect pair of identifiers,’ each identifier would need to be uniquely associated with the nucleic acid analyte of interest.” EX2001, ¶84. I disagree with Dr. Adalsteinsson for multiple reasons.

11. *First*, Dr. Adalsteinsson ignores that claim 1 recites “excluding the sequences of those nucleic acid analytes of interest containing only one identifier *or* an incorrect pair of identifiers.” EX1001 (’200 patent), cl. 1 (emphasis added). A POSA would have understood that the claim requires the exclusion of *either*: (1) sequences of nucleic acid analytes of interest with only one identifier *or* (2) sequences of nucleic acid analytes of interest with an incorrect pair of identifiers, but not necessarily both. As discussed in my previous declaration, a POSA would have understood that sequences containing only one identifier can be identified and excluded, regardless of whether identifier sequences are used for multiple samples. *See* EX1003, ¶¶150-151. This is consistent with *Gloor*, which used a combinatorial dual-indexing method and excluded sequences with only one identifier. *Id.*; EX1005 (*Gloor*), 7-8 (excluding sequences with “incorrect sequence tags” including those with a “complete lack of the left or right end sequence tag”). Notably, in his

deposition, Dr. Adalsteinsson agreed that: (1) the last step of claim 1 is not limited to excluding sequences of nucleic acid analytes of interest containing an incorrect pair of identifiers because it says “or”; and (2) *Gloor* excluded sequences with only one identifier sequence. EX1055 (Adalsteinsson Tr.), 64:21-65:3, 68:7-22, 77:6-12.

12. *Second*, Dr. Adalsteinsson overly limits what constitutes “an incorrect pair of identifiers.” The specification states that an “incorrect pair” refers to “a pair . . . of identifiers not associated with the same template.” EX1001 (’200 patent), 2:23-25. As discussed in my previous declaration, a POSA would have understood that sequences containing an incorrect pair of identifiers (e.g., containing an identifier with truncations, insertions, or erroneous bases) can be identified and excluded, regardless of whether identifier sequences are used for multiple samples. *See* EX1003, ¶¶150-151. This is again consistent with *Gloor*, which used a combinatorial dual-indexing method and excluded sequences with a pair of erroneous identifiers. *Id.*; EX1005 (*Gloor*), 7-8 (excluding sequences with “incorrect sequence tags,” including those containing truncated tags, tags with additional bases, or tags with misincorporated bases).

13. *Third*, Dr. Adalsteinsson ignores that, even with non-unique dual-indexing, a POSA would be able to identify and exclude sequences containing a pair of identifier sequences not initially associated any analytes, provided that some pairs of identifier sequences are not assigned to any sample. To illustrate this, I annotated

the non-unique dual-indexing configuration depicted in my previous declaration (EX1003, ¶78):

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

14. Consider, for example, using this barcode configuration to sequence nucleic acid analytes of interest from 6 samples, assigned to barcode pairs: A-1, B-2, C-1, D-2, E-1, and F-2, respectively (cells highlighted in green). Two identifiers are uniquely associated with each nucleic acid analyte of interest, but some identifiers are used for multiple nucleic acid analytes of interest. Nevertheless, a POSA would be able to identify and exclude sequences containing an “incorrect” pair of identifiers that were not associated with any of the analytes, (e.g., A-2, B-1, C-2, D-1, E-2, F-1, cells highlighted in red). Because only some of the possible combinations of identifiers were originally assigned to samples, the presence of identifier pairs that are “different than any of said distinct pairs” would reflect error. This strategy of using less than all possible combinations of identifiers, referred to

as “sparse” non-unique dual-indexing, illustrates the fallacy of Dr. Adalsteinsson’s argument that “re-using identifiers to create distinct pairs would not allow for identifying an incorrect pair of identifiers to be excluded.” *See* EX2001, ¶84. Notably, other than stating that a sparse indexing strategy would not be his preferred “design choice,” Dr. Adalsteinsson did not deny that one could identify and exclude sequences with a pair of identifiers not originally assigned to any analytes using this strategy. EX1055 (Adalsteinsson Tr.), 56:22-57:18.

15. Consistent with this, it is notable that neither the claims nor the specification specifies the number of samples to be sequenced, the number of identifier sequences to be used, or the ratio between them. *See* EX1001 (’200 patent), cls. 1-6; EX1055 (Adalsteinsson Tr.), 62:2-64:1. The claims thus impose no requirement that all possible combinations of identifier pairs be used and they do not preclude “sparse” non-unique dual-indexing.

16. For these reasons, I disagree with Dr. Adalsteinsson’s implicit claim construction that the term “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest” means that *each* individual identifier sequence is uniquely associated with the analyte of interest. In my opinion, this term should be given its plain and ordinary meaning, i.e., each pair of two or more identifier sequences must be uniquely associated with the nucleic acid analyte of interest. But regardless of how this term

is construed, this limitation was taught in the art and the claims are anticipated and rendered obvious as discussed below.

V. GROUND 1: CLAIMS 1-6 OF THE '200 PATENT ARE ANTICIPATED BY *GLOOR*

A. Dr. Adalsteinsson's Opinion on Claim 1 Is Based Entirely on an Incorrect Claim Construction

17. As explained in my previous declaration, *Gloor* discloses every limitation of claim 1. EX1003, ¶¶136-152. Dr. Adalsteinsson largely does not dispute this, arguing only that *Gloor* does not disclose the limitation of “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest.” EX2001, ¶¶83-85. Dr. Adalsteinsson's analysis, however, is premised entirely on his implicit claim construction that this limitation requires *each* identifier sequence to be uniquely associated with the analyte of interest. As explained above (Section IV), a POSA would have understood that this limitation encompasses using identifier sequences with multiple analytes, provided that no pair of identifier sequences is used with two different analytes.

18. Properly construed, *Gloor* discloses this limitation. As explained in my previous declaration (EX1003, ¶¶141-143), *Gloor* assigned a pair of identifier sequences to each sample. EX1005 (*Gloor*), Abstract. While multiple samples used the same left or right identifier sequence, the *pair* of identifier sequences was unique to that sample. *Id.*; see also *id.*, 6 (Table 2). *Gloor* detected the presence of those

identifier sequences and excluded sequences with a “complete lack of the left or right end sequence tag,” i.e., sequences containing only one identifier. EX1005 (Gloor), 7-8; EX1003, ¶¶139, 151; EX1055 (Adalsteinsson Tr.), 77:6-12. *Gloor* also excluded sequences that contained truncated tags, tags with additional bases, or tags with misincorporated bases, i.e., sequences containing an incorrect pair of identifiers. EX1005 (Gloor), 7-8; EX1003, ¶¶139, 151.

19. Notably, Dr. Adalsteinsson does not dispute that a pair of identifier sequences was uniquely associated with each nucleic acid analyte of interest in *Gloor*. See EX2001, ¶¶83-85. Nor does he dispute that sequences containing only one identifier were excluded in *Gloor*. See *id.* And he ignores that sequences containing incorrect pairs of identifiers, such as sequences with truncated tags, tags with additional bases, or tags with misincorporated bases, were also excluded in *Gloor*. See *id.* As such, under the correct claim construction of the limitation “detecting the presence of two or more identifier sequences that are uniquely associated with the nucleic acid analyte of interest,” *Gloor* discloses every limitation of claim 1 and thus anticipates the claim.

20. In my previous declaration, I opined that *Gloor* discloses every limitation of dependent claims 2-4 and 6 and therefore anticipates those claims. EX1003, ¶¶153-157, 160-161. Dr. Adalsteinsson does not separately dispute my

analysis of those claims. *See* EX2001, ¶¶83-88. Thus, my opinion that those claims are anticipated by *Gloor* remains the same.

B. Dr. Adalsteinsson’s Opinion on Claim 5 Is Contradicted by *Gloor*’s Barcode Sequence Design

21. Dependent claim 5 of the ’200 patent recites that the barcode sequences are “designed to have no homopolymer regions.” EX1001 (’200 patent), cl. 5. Previously, I opined that claim 5 is anticipated by *Gloor* because the barcode sequences used in *Gloor*, as shown in Table 2, have no homopolymer regions. EX1003, ¶¶158-159. Dr. Adalsteinsson does not disagree with my analysis of *Gloor*’s barcode sequences. EX2001, ¶¶86-88. But he disagrees that *Gloor* anticipates claim 5 because *Gloor* “does not instruct a POSA regarding how the sequences in Table 2 were chosen, and specifically does not have any mention of *designing* them to avoid homopolymer regions.” *Id.* (emphasis added).

22. Respectfully, Dr. Adalsteinsson’s argument is nonsensical. As Dr. Adalsteinsson acknowledges (EX2001, ¶87), *Gloor* discusses various considerations that went into “*choos[ing]*” barcode sequences. EX1005 (*Gloor*), 5 (“we *chose* sequence tags to ensure all 4 nucleotides were represented in each of the first four positions of the primers using parameters similar to those in the *barcrawl* program”) (emphasis added). The barcode sequences used in *Gloor* were clearly designed by the authors of *Gloor*, and thus the properties reflected in those sequences are also the products of the authors’ design activities.

23. The fact that **none** of the barcode sequences used in *Gloor* has a homopolymer region further buttresses the conclusion that *Gloor*'s barcode sequences were "designed to have no homopolymer regions." As shown in Table 2, *Gloor* used nine 3-nucleotide barcodes, nine 4-nucleotide barcodes, eight 5-nucleotide barcodes, and eight 6-nucleotide barcodes. EX1005 (*Gloor*), 6 (Table 2). The chance of one n-nucleotide length barcode lacking a homopolymer region simply by chance can be calculated using a simple mathematical formula of $(4 \times 3^{n-1}) / 4^n$. Specifically, "n" represents the number of possible positions in a given barcode (i.e., length). A base for each position can be chosen from four (4) possible bases, A, T, G, C. Once a particular base (e.g., A) is chosen for a specific position, the next position can only have a base selected from the three (3) remaining bases (i.e., T, G, C) to avoid a homopolymeric run.

24. Using this formula, the chance of one 3-nucleotide barcode having no homopolymer region is 56% ($=36/64$). But the chance of **all nine** 3-nucleotide barcodes having no homopolymer region is orders of magnitude lower, only about 0.56% ($= (36/64)^9$). Similarly, the chance of one 4-nucleotide barcode having no homopolymer region is about 42% ($=108/256$). But the chance of **all nine** 4-nucleotide barcodes having no homopolymer region is only about 0.042% ($= (108/256)^9$). Similarly, the chance of all eight 5-nucleotide barcodes having no homopolymer region is about 0.01% ($= (324/1,024)^8$) and the chance of **all eight** 6-

nucleotide barcodes having no homopolymer region is only about 0.001% (= $(972/4,096)^8$). The chance of *all* the barcodes in *Gloor* having no homopolymer region simply by chance is incredibly small, about 2.4×10^{-15} , which is about 2 in 1 quadrillion (or about 1 million times lower odds than winning the Powerball lottery). In my opinion, this compels the conclusion that the barcodes in *Gloor* were “designed” to have no homopolymer region, which is consistent with a POSA’s understanding in view of the long history of barcode design. *Gloor* discloses each and every limitation of claim 5 and thus anticipates the claim.

VI. GROUND 2: CLAIMS 1-6 OF THE ’200 PATENT ARE ANTICIPATED BY *PARAMESWARAN*

A. Dr. Adalsteinsson’s Opinion on Claim 1 Ignores Important Disclosures of *Parameswaran*

25. As I explained in my previous declaration, *Parameswaran* discloses every limitation of claim 1, including the last limitation: “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers from sequence analysis of the multiplex sequencing reaction.” EX1003, ¶¶163-181. Specifically, *Parameswaran* classifies the various barcode motif combinations into the following categories:

Sequences that have a perfect match to only one of the expected motifs have been further subdivided into three classes based on whether the motif at the other end has: (i) perfect identity to an unexpected motif, resulting in an unassigned 5'–3' motif combination (perfect, unexpected occurrence), (ii) a high-scoring partial match to the expected motif for that library (imperfect, expected occurrence) and (iii) no strong match to any barcode motif (imperfect, ambiguous occurrence). Perfect, 'unexpected'

EX1004 (Parameswaran), 7. Using a unique dual-indexing method, *Parameswaran* identified sequences with mismatched barcode motifs, i.e., a pair of barcode motifs where each barcode motif would assign the sequence to a different sample, and which therefore should not have been found in the same sequence, classifying them as “perfect, ***unexpected***.” *Id.* (emphasis added). *Parameswaran* refers to these sequences as “false-discover[ies],” constituting “misassignment[s]” of sequences to incorrect samples. *Id.*, 7-8. In contrast, *Parameswaran* validated sequences containing “both 5’ and 3’ barcode motifs” (i.e., sequences with a correct pair of identifiers) through placement “into the correct bin.” *Id.*, 8. Thus, *Parameswaran* discloses every limitation of claim 1, including the “validating” limitation.

26. In response, Dr. Adalsteinsson does not dispute that *Parameswaran* identified sequences with mismatched barcode motifs, i.e., an incorrect pair of identifier sequences. See EX2001, ¶¶89-91. But he argues that *Parameswaran* did not exclude them. For example, from the sentence in *Parameswaran* 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,” Dr. Adalsteinsson infers that the only sequences *Parameswaran* “excluded” were those without a match to any barcode. *Id.* (citing EX1004 (*Parameswaran*), 8).

27. 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.

28. *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	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	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	40
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	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	26	20	48	29	64	46	42	36	78	59	60	47	58	35	52	26	77	27	70	59	29		

Figure 1. *Parameswaran's* Supplemental Table 5D (Original)

29. 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 the 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	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	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 2. *Parameswaran's* Supplemental Table 5D (Annotated)

30. 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.

31. 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.*

32. 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.

33. *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~~9 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 excluded from the sequencing analysis because they could not be assigned to any one sample.

34. In sum, contrary to Dr. Adalsteinsson's assertion, and as explained in my prior declaration and above, *Parameswaran* discloses "excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers" as recited in claim 1 of the '200 patent. Dr. Adalsteinsson does not dispute my opinion that *Parameswaran* disclosed all other limitations in claim 1. See EX2001, ¶¶89-91. As such, *Parameswaran* discloses every limitation of claim 1, and it therefore anticipates the claim.

35. In my previous declaration, I opined that *Parameswaran* also discloses every limitation of dependent claims 2-4 and therefore anticipates those claims. EX1003, ¶¶182-186. Dr. Adalsteinsson does not separately dispute my analysis of those claims. See EX2001, ¶¶89-100. Thus, my opinion that those claims are anticipated by *Parameswaran* remains the same.

B. Dr. Adalsteinsson's Opinion on Claim 5 Ignores *Parameswaran's* General Teaching Regarding Barcode Design

36. In my earlier declaration, I opined that *Parameswaran* discloses the limitation in dependent claim 5 that "barcode sequences are designed to have no homopolymer regions." EX1003, ¶¶187-188. Dr. Adalsteinsson disagrees, contending that "every one of *Parameswaran's* barcodes contains a homopolymer region." EX2001, ¶96. However, by focusing on *Parameswaran's* specific examples, Dr. Adalsteinsson minimizes *Parameswaran's* general teaching that barcodes should be designed to "avoid homopolymeric runs." EX1004

(Parameswaran), 4 (“barcode design [] facilitated unambiguous assignment, with care taken to avoid homopolymeric runs”); *id.* (“[t]his design avoids significant problems due to error rate at homopolymeric runs of nucleotides”). Regardless of whether its specific barcodes have homopolymer regions, *Parameswaran* overall discloses “barcode sequences” that are “design[ed]” to have no homopolymer regions. *Id.*

37. A POSA would further have had access to preexisting knowledge in barcode design. Notably, the portion of the ’200 patent that discusses avoiding homopolymer regions was copied essentially verbatim from a prior patent EX1037 (Steinman), item (43), 1, 2:3-4. In the context of a method of sequencing a plurality of nucleic acids from different samples, that patent disclosed that the barcodes “are designed to have minimal or no homopolymer regions, i.e., 2 or more of the same base in a row.” *Id.*, 5:2-5. As shown below, this is nearly identical to how the ’200 patent defines the term “homopolymer region.” EX1001 (’200 patent), 9:5-11. *Parameswaran*, as interpreted from the perspective of a POSA, discloses designing barcodes with no homopolymer region.

EX1001 (’200 Patent), 9:5-11	EX1037 (Steinman), 5:2-5
The barcode sequence generally includes certain features that make the	The oligonucleotide sequence generally includes certain features that

sequence useful in sequencing reactions. For example the barcode sequences can be designed to have minimal or no homopolymer regions, i.e., 2 or more of the same base in a row such as AA or CCC, within the barcode sequence.	make the sequence useful in sequencing reactions. For example the oligonucleotide sequences are designed to have minimal or no homopolymer regions, i.e., 2 or more of the same base in a row such as AA or CCC, within the unique portion of the oligonucleotide sequence.
---	---

C. Dr. Adalsteinsson’s Opinion on Claim 6 Adds an “Attachment” Limitation Not Required by the Claim

38. As previously explained, *Parameswaran* sequenced analytes of interest using the Roche 454 platform, which involves attaching a nucleic acid analyte of interest to beads and placing those beads in wells of a “picotiter plate” (a flow cell). EX1003, ¶190. As such, I opined that *Parameswaran* anticipates claim 6, which recites that “the nucleic acid analyte of interest is on the surface of a flow cell.”

39. According to Dr. Adalsteinsson, *Parameswaran* does not anticipate claim 6 because “the beads [in the Roche 454 platform used by *Parameswaran*] are free-floating and ***not attached*** to the flow cell.” EX2001, ¶¶97-100 (emphasis added). In other words, Dr. Adalsteinsson adds an ***attachment*** requirement to the claim that is not required by claim 6. As explained in my previous declaration,

claim 6 does *not* require the analyte of interest to be attached to the surface of a flow cell. EX1003, ¶190. It only requires the analyte to be “on the surface of a flow cell,” which is how the Roche 454 platform works. *See, e.g.*, EX1010 (Kircher), 3-4 (“The sequencing plate [of Roche 454] has about two million wells — each of them able to accommodate exactly one 28-µm diameter bead covered with single-stranded copies of the sequence to be determined.”). Further, the beads containing DNA in the Roche 454 system were not “free-floating”—it would defeat the ability of the Roche 454 sequencer to sequence if the beads could change position, floating freely from well to well. Instead, the beads containing the DNA were held firmly in place at the surface of the well by a further set of smaller packing beads. Dr. Adalsteinsson does not dispute that the Roche 454 platform disclosed in *Parameswaran* involves attaching a nucleic acid analyte of interest to beads and placing those beads in wells of a flow cell. *See* EX2001, ¶¶97-100. Nor does he dispute that the wells in which the beads are deposited are located on the surface of the flow cell. *See id.* As such, *Parameswaran* discloses each and every limitation of claim 6 and thus anticipates the claim.

VII. GROUND 3: CLAIMS 1-6 OF THE '200 PATENT WOULD HAVE BEEN OBVIOUS OVER *PARAMESWARAN* AND *GLOOR*

A. Dr. Adalsteinsson's Obviousness Opinion Focuses on Unclaimed Features

40. The claims of the '200 patent recite: “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and *excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers* from sequence analysis of the multiplex sequencing reaction” (emphasis added). As discussed above, *Gloor* disclosed excluding sequences containing only one identifier. *See supra* §§ IV, V.A. *Gloor* also disclosed excluding sequences with an incorrect pair of identifiers. *Id.* *Parameswaran* disclosed excluding sequences with an incorrect pair of identifiers not initially assigned to any of the analytes. *See supra* § VI.A. As I explained in my previous declaration, a POSA would have been motivated to exclude these erroneous sequences during multiplex sequencing. EX1003, ¶203. A POSA would have had a reasonable expectation of success in doing so because *Gloor* and *Parameswaran* had already successfully done so. *Id.*, ¶209. Prior art studies routinely discarded PCR-induced artifacts to improve sequencing accuracy, demonstrating that a POSA would have been motivated, with a reasonable expectation of success, to discard erroneous 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”). For at least this reason, the claims of the ’200 patent would have been obvious to a POSA in December 2010.

41. Notably, Dr. Adalsteinsson does not dispute that *Gloor* excluded erroneous sequences, such as sequences with only one identifier or sequences with an incorrect pair of identifiers. *See* EX2001, § XIII. Instead, a substantial portion of Dr. Adalsteinsson’s obviousness opinion focuses specifically on formation of chimeric sequences, including by cross-over error. *See id.* For example, Dr. Adalsteinsson’s declaration uses the term “crossover” nine times, the term “cross-over” twenty-four times, and the term “chimeric” ninety-one times. None of these terms are used in the claims of the ’200 patent, which are not limited to excluding chimeric sequences formed by cross-over error. Instead, they are more broadly directed to validating the sequence of a nucleic acid analyte of interest by excluding those sequences containing only one identifier sequence or an incorrect pair of identifier sequences. *See supra* § IV.

42. Dr. Adalsteinsson nonetheless argues (among other arguments) that the claimed methods could not have been obvious because there was “a general lack of awareness of chimeric sequences formed by the cross-over between two samples in

NGS machines” such as “flow-cell based NGS.” *E.g.*, EX2001, ¶126. I respectfully disagree with his opinion, and address each of his arguments below.

B. Claims 1-4 and 6 Would Have Been Obvious over *Parameswaran* and *Gloor*

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

43. In my previous declaration, I explained that the formation of chimeric molecules was a known problem for NGS platforms as of December 2010. EX1003, ¶¶69-71. Dr. Adalsteinsson disagrees, contending that there was a “general lack of awareness of cross-over error in NGS machines” as of December 2010. EX2001, ¶¶115-126; *see also id.*, ¶¶36, 42, 137-138. I respectfully disagree.

a. Chimeric Molecules, Including Those “Formed by the Cross-Over Between Two Samples,” Were a Known Problem

44. Contrary to Dr. Adalsteinsson’s characterization, and as explained in my prior declaration, chimeric molecules, including those “formed by the cross-over between two samples,” were a known problem for NGS platforms such as Illumina sequencers as of December 2010.

45. *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, ¶70. 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.” *See* EX1003, ¶62.

46. 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, ¶¶63-64. 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, ¶¶63-64.

47. 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.”).

48. 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.

49. 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.

50. *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.*

51. 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* §§ VII.B.2-5.

52. *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.

53. *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.*

54. 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).

55. *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”).

56. 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.*

57. *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.

58. 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").

59. 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.

60. 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, ¶31; 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

61. 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 that could be addressed by using [unique dual-indexing].” EX2001, ¶116; *see id.*, ¶¶115-123 (Dr. Adalsteinsson discussing EX1017 (Lahr), EX1018 (Meyerhans), EX1019 (Paabo), EX1020 (Odelberg), EX1033 (Dai), and EX1034 (Creer)).

62. 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, ¶70. 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.

63. 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.

64. 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, ¶¶36, 42 (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 and would “produce chimeras” that could be identified and excluded by unique dual-indexing. See EX2001, ¶116.

65. Dr. Adalsteinsson cites additional articles that he contends show that “cross-over error formed by chimeric sequences formed between samples for flow-cell based NGS was not thought to be a problem” as of December 2010. EX2001, ¶¶36, 42, 124, 137-138 (citing EXS2003-2007). Again, I disagree with Dr. Adalsteinsson’s interpretation of this evidence.

66. EX2005: Citing EX2005, Dr. Adalsteinsson states that “literature at the time of the invention explained that cross-over error formed by chimeric sequences formed between samples for flow-cell based NGS was not thought to be a problem.” EX2001, ¶124. 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 formed by chimeric sequences formed between samples for flow-cell based NGS was not thought to be a problem.” *See* EX2001, ¶124.

67. 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.

68. 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*.

69. 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* § VII.B.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, 2010.¹

70. EXS2003-2004: Relying on EX2003 and EX2004, Dr. Adalsteinsson asserts that “the industry did not appreciate the cross-over error caused by chimeric molecules formed during cluster amplification.” EX2001, ¶¶36, 42, 138. Setting aside that both documents published after December 2010, they do not support Dr. Adalsteinsson’s characterization of the art as of December 2010.

71. 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

¹ 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 formed by chimeric sequences formed between samples for flow-cell based NGS was not thought to be a problem” (EX2001, ¶124) 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.

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, ¶70) 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.

72. EXS2003, 2006-2007: Dr. Adalsteinsson asserts that "Illumina itself provided no indication that it was aware of cross-over error due to chimeric sequences, or that it was an issue on Illumina sequencers." EX2001, ¶¶137-138. 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 due to chimeric sequences 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).

73. In sum, numerous prior-art studies confirm that chimeric molecules, including those “formed by the cross-over between two samples,” 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

74. Dr. Adalsteinsson asserts that a POSA would not have been motivated to use *Parameswaran*’s unique dual-indexing method to exclude “chimeric molecules leading to cross-over error.” EX2001, ¶¶126-136. But a POSA would have recognized *Parameswaran*’s method as a suitable solution to the known problem of chimeric molecules between different samples.

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

75. 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* § VII.B.1. A POSA would have been motivated to identify and exclude sequences from chimeric molecules formed from different samples regardless of when they arise because these chimeric artifacts do not

represent the sequences as originally obtained from the biological samples. *See supra* § VII.B.1; EX1003, ¶202.

76. In December 2010, software was often used to identify these artificial chimeras. *See* EX1003, ¶203. 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.

77. *Parameswaran* offered a solution to this problem. *See* EX1003, ¶203. Using a unique dual-indexing method, *Parameswaran* demonstrated the ability to identify chimeric sequences created during multiplex sequencing based on mismatched barcode motifs—i.e., barcode motifs that should not appear together because they originated from different samples. *Id.*, ¶¶203, 163-167; *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]”).

78. Notably, *Parameswaran*’s unique dual-indexing methodology was not limited to identifying chimeric sequences formed at a specific stage of multiplex sequencing. Instead, it could detect inter-sample chimeric sequences that occurred any time after barcoding, providing even more power to the technology. As illustrated below, using *Parameswaran*’s barcoding strategy, sequences from chimeric molecules arising before (Figure 3) and after (Figure 4) the samples were added to the Roche 454 or Illumina sequencer would contain mismatched barcodes. Dr. Adalsteinsson also acknowledged that sequences with mismatched barcodes can be generated at various stages, including during PCR amplification of pooled

samples prior to cluster amplification and during cluster amplification. EX1055
(Adalsteinsson Tr.), 94:21-97:10.

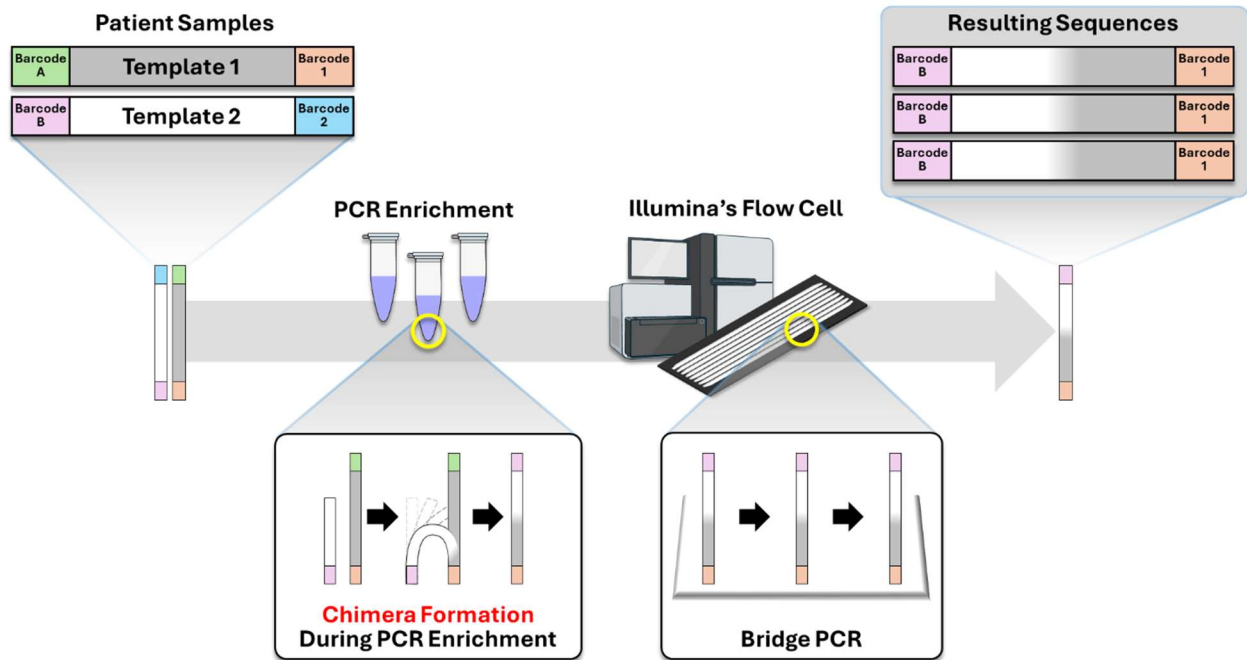


Figure 3. Sequences of Chimeric Molecules Created Before Amplification “on the Surface of the Flow Cell”

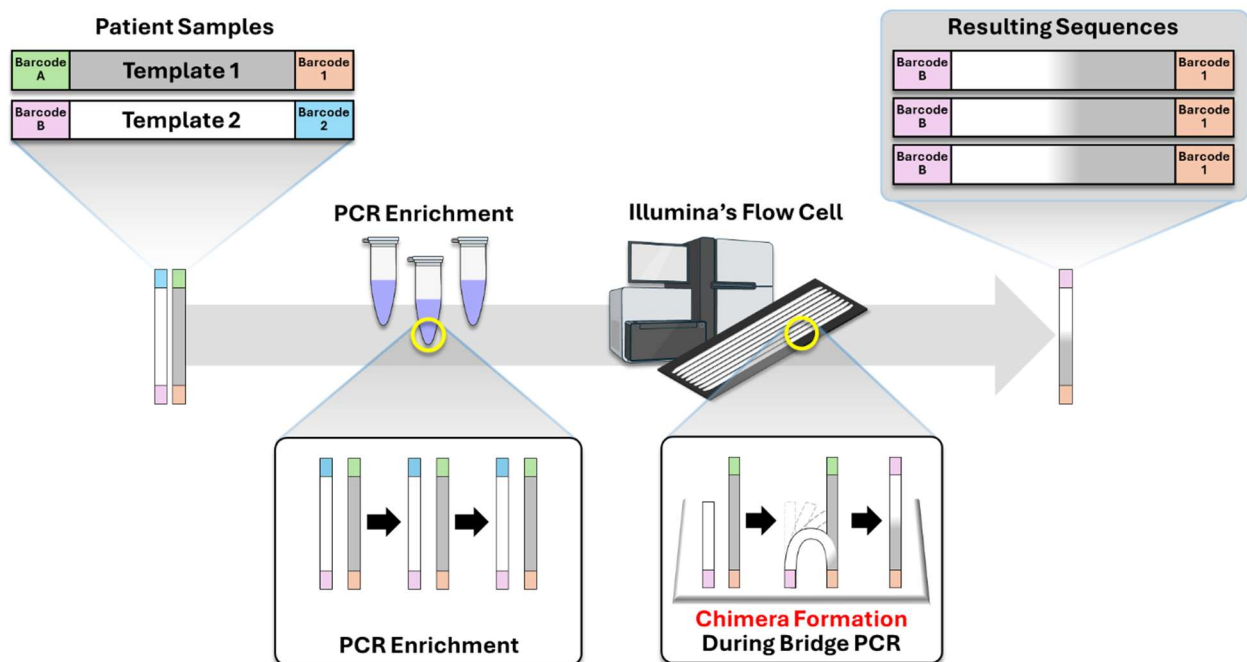


Figure 4. Sequences of Chimeric Molecules Created During Amplification “on the Surface of the Flow Cell”

79. In the context of the claims of the '200 patent, sequences of chimeric molecules made of templates from different samples “on the surface of beads” or “on the surface of the flow cell” are indistinguishable from those created at other stages during multiplex sequencing—both contain pairs of mismatched identifiers that should not be found together. As illustrated below, excluding sequences with mismatched identifiers using *Parameswaran*’s unique dual-indexing method would exclude all of these chimeric sequences, regardless of when and how they formed during multiplex sequencing.

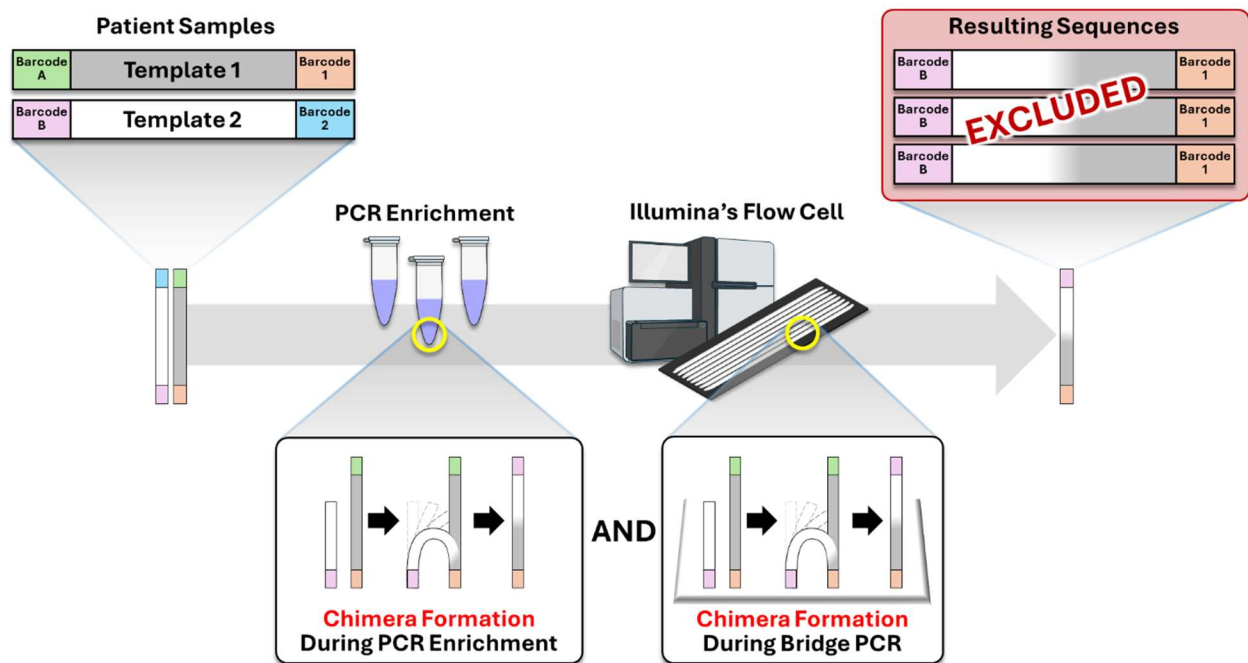


Figure 5. *Parameswaran*’s Method Excludes Chimeric Sequences Formed at Various Stages During Multiplex Sequencing

80. As such, using *Parameswaran*’s method, a POSA performing multiplex sequencing on an Illumina platform would have been able to identify and exclude

sequences from chimeric molecules made of templates from different samples introduced during multiplex sequencing.

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

81. 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, ¶127. I disagree. As explained above, a POSA in December 2010 would have understood that chimeric molecules would emerge at multiple stages during multiplex sequencing, 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* § VII.B.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.² By 2010, a POSA would have readily understood the

² Notably, one of the causes identified in *Parameswaran* for sequences with mismatched barcode motifs was “cross-contamination of template DNAs in the

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.”).

82. Dr. Adalsteinsson also speculates that “the mismatched error-free barcode motifs observed in *Parameswaran* were not a result of chimeric molecules leading to cross-over.” EX2001, ¶126. According to Dr. Adalsteinsson, this is because a POSA would have understood that “the creation of chimeric molecules 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.

83. As Dr. Adalsteinsson noted, “six of the forward barcodes are exact reverse complements of six reverse barcodes” in *Parameswaran*. EX2001, ¶126;

second round of PCR.” EX1004 (*Parameswaran*), 7. A POSA would have understood that this is equivalent to “cross-contamination of template DNAs” during 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.

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 sequencing reads generated from 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 (Parameswaran), 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.

84. As *Parameswaran* 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, ¶¶86, 164. To address this ambiguity, *Parameswaran* combined the 10-nt barcodes with three bases from the adjoining linker (forming a “barcode motif”). EX1004 (Parameswaran), 7; EX1003, ¶¶86, 164. This made each barcode motif unique to each sample, because neither the sequence nor its reverse complement was repeated across different samples. *See* EX1004 (Parameswaran), 16 (Suppl. Table 1B).

Importantly, while this configuration reduced the number of sequences with conflicting barcodes, it did *not* eliminate them.

85. 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.

86. 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, ¶130. 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.

87. 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.

88. 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, ¶¶131-132. 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* § VII.B.2.a; 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.

89. Citing EX2009, Dr. Adalsteinsson argues that, because “a 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 multiplex sequencing. EX2001, ¶¶112-114. 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.

90. 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, ¶¶134-135 (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.

91. 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.

92. 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.”).

93. 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 Reasons for *Not* Combining *Parameswaran* and *Gloor* Lack Scientific Merit

94. In my previous declaration, I identified multiple reasons why a POSA would have been motivated to combine *Parameswaran* and *Gloor*. See EX1003, ¶¶202-208. Dr. Adalsteinsson identifies purported reasons why a POSA *would not* have combined these references. EX2001, ¶¶102-108, 140. As discussed below, I disagree with Dr. Adalsteinsson’s opinions.

95. *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, ¶¶103-104. 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, ¶¶70-72. Moreover, the formation of chimeric molecules was recognized as an inherent risk in multiplex sequencing workflows, irrespective of the specific platform employed. *See supra* § VII.B.1; EX1003, ¶¶70-72, 61-64. Indeed, *Gloor* explicitly documents the occurrence of chimeras following multiplexed sequencing conducted on an Illumina instrument. *See* EX1005 (*Gloor*), 8, 22.

96. 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.*

97. *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.*

98. 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.

99. An inventor of the '200 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).

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

101. *Second*, Dr. Adalsteinsson asserts that because *Parameswaran*'s barcodes are "tailored" for pyrosequencing, its approach should be limited to

pyrosequencing. EX2001, ¶¶102-104. 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, ¶¶210-213; 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, ¶212. 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.

102. The ’200 patent further confirms that barcode design was not a barrier. EX1003, ¶213. The ’200 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 (’200 patent), 3:44-48, 8:65-9:3 (citing prior art approaches for “designing sets of barcode sequences”),

20:40-44 (“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.”)

103. I also disagree with Dr. Adalsteinsson’s related argument that *Gloor* “cautioned” against applying pyrosequencing approaches to Illumina sequencers. EX2001, ¶105. 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, ¶¶95-98. A POSA, familiar with these distinctions, would have understood *Gloor* as offering clear guidance for adapting *Parameswaran*’s strategy to Illumina sequencing.

104. *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, ¶¶106-107. I disagree with Dr. Adalsteinsson’s overgeneralized assessment.

105. 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* §§ VII.B.1-2. 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, ¶203. 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.

106. *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, ¶206. 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, ¶140.

107. 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.

108. Dr. Adalsteinsson cannot dispute that Illumina sequencers were one of the major NGS platforms available in December 2010. EX1003, ¶52. 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.

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

109. Dr. Adalsteinsson criticizes me for “only focus[ing] on the costs and efficiencies of the sequencers themselves and ignore[ing] extra costs and work associated with *Parameswaran’s* unique dual indexes.” EX2001, ¶¶109-112; *see also id.*, ¶107. 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.*, ¶111.

110. 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.

111. 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* §§ VII.B.1-3. Reflecting this, researchers (including an inventor of the ’200 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* § VII.B.3. Dr. Adalsteinsson disregards these factors, which would have strongly motivated a POSA to apply *Parameswaran’s* unique dual-indexing to the Illumina sequencer in *Gloor*.

112. 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.

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

113. Dr. Adalsteinsson contends that “a POSA in 2010 would not have understood the prevalence of chimeric molecules forming 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.” EX2001, ¶141. He further contends that “[b]ecause a POSA would not have been aware of chimeric sequences between samples or the problems caused by such sequences, a POSA would not have been aware that the exclusion criteria presented in the patent would indicate an invalid sequence.” *Id.* Like his motivation arguments, Dr. Adalsteinsson's expectation of success arguments are premised on his incorrect characterization of the state of the art.

114. As I discuss above, by December 2010, it was known that chimeric sequences could form at various stages during the sequencing performed on NGS platforms. *See supra* § VII.B.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* § VII.B.1, there was no scientific basis for asserting that chimeric molecules would not form on NGS platforms (as opposed to traditional PCR) or among templates from different samples (as opposed to a single sample). *Id.* Tellingly, Dr. Adalsteinsson does not appear to dispute that *Parameswaran*'s or *Gloor*'s methods would validate the sequence of a nucleic acid analyte of interest in a multiplex sequencing. *See* EX2001, ¶141.

115. Dr. Adalsteinsson also asserts that a POSA would not have expected “any utility to using UDI over CDI on Illumina’s sequencers.” EX2001, ¶141. 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* §§ VII.B.1-4. 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 excluding chimeric sequences that could lead to critical sample misassignment. *Id.*

116. 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 claim

1 of the '200 patent. To a POSA in December 2010, claim 1 of the '200 patent would have been obvious over *Parameswaran* and *Gloor*.

117. As explained in my previous declaration, the combination of *Parameswaran* and *Gloor* also renders dependent claims 2-4 and 6 obvious. EX1003, ¶¶216-218, 220-221. Dr. Adalsteinsson does not separately dispute my analysis of those claims. Thus, my opinion that those claims would have been obvious over *Parameswaran* and *Gloor* remains the same.

C. Claim 5 Would Have Been Obvious Over *Parameswaran* and *Gloor*

118. As discussed above, dependent claim 5 of the '200 patent requires that the barcode sequences be “designed to have no homopolymer regions.” EX1001 ('200 patent), cl. 5. Solely relying on his anticipation analyses, Dr. Adalsteinsson contends that “[n]either *Parameswaran* nor *Gloor* discloses this limitation.” EX2001, ¶142. Not so. As explained above, both *Parameswaran* and *Gloor* disclose the design of barcodes having no homopolymer regions. *See supra* §§ V.B, VI.B.

119. Further, a POSA would have been motivated to design barcode sequences that have no homopolymer regions based on *Parameswaran*'s teaching to avoid homopolymer regions in barcode sequences. EX1004 (*Parameswaran*), 4 (“barcode design [] facilitated unambiguous assignment, with care taken to avoid homopolymeric runs”); *id.* (“[t]his design avoids significant problems due to error rate at homopolymeric runs of nucleotides”); EX1003, ¶188. A POSA would have

had a reasonable expectation of success in doing so, as was successfully done in *Gloor*. EX1005 (*Gloor*), 6 (Table 2) (all barcode sequences lacked homopolymer regions); EX1003, ¶159. Thus, claim 5 would have been obvious in view of *Parameswaran* and *Gloor*.

VIII. GROUND 4: CLAIMS 1-6 OF THE '200 PATENT WOULD HAVE BEEN OBVIOUS OVER *PARAMESWARAN* IN VIEW OF *BINLADEN*

A. Claims 1-4 Would Have Been Obvious Over *Parameswaran* and *Binladen*

120. With respect to the combination of *Parameswaran* and *Binladen*, Dr. Adalsteinsson raises two issues: (1) the references allegedly do not teach or suggest the last limitation where both identifiers are analyzed and sequences with only one identifier or an incorrect pair of identifiers are excluded and (2) there was purportedly no motivation to combine these references. EX2001, ¶¶143-145, 152-154. As explained below, I disagree with both.

1. Dr. Adalsteinsson's Opinion on the "Excluding" Step Is Incorrect

121. As with his anticipation analyses, Dr. Adalsteinsson focuses on a single limitation of claim 1: "analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers." EX2001, ¶¶143-145. He first asserts that *Parameswaran* does not disclose this limitation. EX2001, ¶143. This is incorrect for the reasons set forth above. *See supra* §§ VI.A, VII.B.2.

122. Dr. Adalsteinsson then asserts that *Binladen* does not disclose this limitation. EX2001, ¶¶144-145. Specifically, he theorizes that, because the length of reads in *Binladen* was approximately 120 bases and the full sequence length of the template DNA analytes was 133-141 bases, “Binladen cannot sequence all of the first primer tag, the template DNA, and the second primer tag in its sequencing reads.” *Id.*, ¶144. Based on this theory, Dr. Adalsteinsson states that “Binladen includes sequences with only one identifier sequence” and “could not observe incorrect pairs of identifiers because it only analyzed sequences with one primer sequence.” *Id.*, ¶145.

123. As an initial matter, Dr. Adalsteinsson confuses the *sequence* of an analyte of interest with individual *reads* of the sequence. Claim 1 of the ’200 patent does not require that the two identifiers of a given analyte of interest must be sequenced in a single read or analyzed from a single read. In other words, a POSA would have understood that the claim allows for sequencing, detecting, and analyzing two identifiers associated with an analyte of interest from different reads.

124. Properly understood, *Binladen* discloses a method for “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers.” As explained in my previous declaration, *Binladen* reports that each analyte generated two sets of reads—each

containing either the forward or reverse tag—collectively constituting the entire sequence of the analyte. EX1029 (Binladen), 3 (“The implication of this is that approximately 50% of each PCR product will be sequenced from the orientation of the forward primer, and 50% from the orientation of the reverse primer.”); EX1003, ¶237. *Binladen* also explains that both reads, each containing a forward or reverse tag, were analyzed by comparing them against the reference sequence using “direct and reverse complementation.” EX1029 (Binladen), 5; EX1003, ¶241. As such, *Binladen* discloses “analyzing both identifiers,” as recited in claim 1.

125. A POSA would have also understood that *Binladen* discloses a method for “validating the sequence of the nucleic acid analyte of interest by analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers.” As explained above, *Binladen* analyzed both identifiers of the sequence of the analyte of interest. *Binladen* then “discarded all sequence reads without an exact match to any of the primers used in the studies,” including all sequences that “lacked the 5’ tag sequence.” EX1029 (Binladen), 5-6. Because each sequence of the analyte in *Binladen* contains two **reads** (forward and reverse), a sequence containing only one identifier necessarily has one of its two reads lacking an identifier. Thus, as explained in my previous declaration, by excluding sequence reads lacking an

identifier, *Binladen* necessarily excluded the sequences of analytes containing only one identifier identifiers. EX1003, ¶241.

126. For these reasons, a POSA would have understood that *Binladen* discloses the limitation “analyzing both identifiers and excluding the sequences of those nucleic acid analytes of interest containing only one identifier or an incorrect pair of identifiers” in claim 1.

2. Dr. Adalsteinsson’s Opinion Regarding Motivation to Combine *Parameswaran* and *Binladen* Reads These References in Isolation

127. According to Dr. Adalsteinsson, a “POSA would not have been motivated to use the dual indexing methods of both *Parameswaran* and *Binladen*” because *Parameswaran* commented on the short length of barcodes used in *Binladen* and *Parameswaran* considered its method to be “superior” to *Binladen*’s method. EX2001, ¶¶152-154. Contrary to Dr. Adalsteinsson’s assertion, a POSA would have been motivated, not dissuaded, to combine the teachings of *Parameswaran* and *Binladen*.

128. A POSA in December 2010 reading these references would have appreciated strengths and weaknesses of the dual-indexing methods disclosed in *Parameswaran* and *Binladen* and understood that the combined teachings of these references could further improve sequencing accuracy. *Parameswaran*, for example, excluded sequences with an incorrect pair of identifiers, but it did not

exclude sequences with only one identifier “[f]or our purposes.” *See supra* § VI.A; EX1004 (Parameswaran), 8. In contrast, *Binladen* excluded sequences with only one identifier, recognizing those sequences as “errors.” EX1029 (Binladen), 5-6; *see supra* § VIII.A.1. To enhance the accuracy of multiplex sequencing, a POSA would have been motivated to adopt *Parameswaran*’s unique dual-indexing method, including using longer barcodes than those disclosed in *Binladen*. A POSA would have been further motivated to ensure sequence fidelity by applying the validation methods utilized in *Parameswaran* and *Binladen*, i.e., excluding sequences with one identifier, as in *Binladen*, or an incorrect pair of identifiers, as in *Parameswaran*. In my opinion, these references complement each other, rather than teach away as Dr. Adalsteinsson asserts.

129. Dr. Adalsteinsson does not dispute that a POSA would have had a reasonable expectation of success in combining *Parameswaran* and *Binladen*. Thus, for the reasons explained above, a POSA would have found that claim 1 is obvious over the combination of *Parameswaran* and *Binladen*.

130. As explained in my previous declaration, the combination of *Parameswaran* and *Binladen* also renders dependent claims 2-4 obvious. EX1003, ¶¶247-249. Dr. Adalsteinsson does not separately dispute my analysis regarding those claims. Thus, my opinion that those claims would have been obvious over *Parameswaran* and *Binladen* remains the same.

B. Claims 5 Would Have Been Obvious Over *Parameswaran* and *Binladen*

131. In my previous declaration, I opined that claim 5 would have been obvious over *Parameswaran* and *Binladen* because a POSA would have been motivated to use barcode sequences without homopolymer regions, as taught in *Parameswaran*, with a reasonable expectation of successfully doing so, as had already been done in *Binladen*. EX1003, ¶250. Referring to his anticipation analyses, Dr. Adalsteinsson contends that *Parameswaran* does not disclose barcode sequences designed to have no homopolymer regions. EX2001, ¶146. For the reasons discussed above, I disagree. *See supra* § VI.B.

132. I also disagree with Dr. Adalsteinsson's argument that *Binladen* does not disclose this limitation. EX2001, ¶¶146-150. Focusing on dinucleotide tags in *Binladen*, some of which included repeat bases, Dr. Adalsteinsson alleges that I “[did] not address Binladen’s dinucleotide tags but referenc[ed] its tetranucleotide tags.” *Id.* Dr. Adalsteinsson further contends that: (1) a “POSA would be wary of ignoring and discarding Binladen’s dinucleotide tags” and (2) *Binladen* “does not disclose that its barcodes are designed to have no homopolymer regions.” *Id.*

133. *First*, I did not ignore *Binladen*’s dinucleotide tags. I specifically discussed two types of barcodes used in *Binladen*: (1) a first set containing dinucleotide tags and (2) a second set containing tetranucleotide tags. EX1003, ¶105. I also discussed *Binladen*’s disclosure comparing these two types of

barcodes—while the tetranucleotide tags lowered PCR efficiency, “overall the rate of sequence miss-assignment for these primers was lower than for the dinucleotide tags.” *Id.*, ¶108 (citing EX1029 (Binladen), 7, 8).

134. *Second*, a POSA would not have been “wary of ignoring and discarding Binladen’s dinucleotide tags.” As summarized in my pervious declaration, *Binladen* discussed pros and cons of these sets of barcodes, including “the apparent reduced miss-assignment rates of the tetranucleotide tags.” EX1029 (Binladen), 8; EX1003, ¶105. To a POSA, it would have been obvious to choose appropriate barcode sequences depending on the sequencing purpose, and if “the apparent reduced miss-assignment rates” were desired, it would have been obvious for a POSA to consider longer tags. Dr. Adalsteinsson does not dispute that the tetranucleotide tags in *Binladen* have no homopolymer regions. *See* EX2001, ¶147 (Dr. Adalsteinsson reproducing barcode sequences used in *Binladen*, including the tetranucleotide tags lacking homopolymer regions).

135. *Third*, Dr. Adalsteinsson’s “did not design” arguments, which he already raised against *Gloor*, are unpersuasive for the reasons discussed above. *See supra* § V.B. All of the tetranucleotide tags in *Binladen* have no homopolymer region. A POSA would have understood that the authors of *Binladen* specifically designed those tags and that those tags avoided homopolymer regions. *Id.* Further, even if *Binladen* does not expressly mention designing barcode sequences without a

homopolymer region, this feature would have been obvious in view of *Parameswaran*'s express recommendation to "avoid homopolymeric runs." EX1004 (*Parameswaran*), 4.

C. Claim 6 Would Have Been Obvious Over *Parameswaran* and *Binladen*

136. Claim 6 of the '200 patent recites that "the nucleic acid analyte of interest is on the surface of a flow cell." EX1001 ('200 patent), cl. 6. As with *Parameswaran*, Dr. Adalsteinsson contends that "Binladen also does not disclose this limitation for the same reasons because . . . Binladen uses the same pyrosequencing that *Parameswaran* used." EX2001, ¶151. As discussed above, claim 6 specifies the location of the nucleic acid analyte of interest to be "on the surface of a flow cell." *See supra* § VI.C. This feature would have been obvious over *Parameswaran* and *Binladen*, both of which used the Roche 454 platform where beads with nucleic acid analytes are sequenced on the surface of a flow cell. *Id.*

IX. GROUND 5: CLAIM 6 OF THE '200 PATENT WOULD HAVE BEEN OBVIOUS OVER *PARAMESWARAN*, *BINLADEN*, AND *BENTLEY*

137. As explained in my previous declaration and above, claim 6 does not require attaching the analyte to the surface of a flow cell; it only requires the nucleic acid analyte of interest to be "on the surface of a flow cell." *See supra* § VI.C. But even if the claim required attaching analytes to the surface of a flow cell, *Bentley*

discloses this limitation, and a POSA would have been motivated to combine *Bentley* with *Parameswaran* and *Binladen*, with a reasonable expectation of success. EX1003, ¶¶253-257.

138. Dr. Adalsteinsson does not dispute that *Bentley* discloses attaching analytes to the surface of a flow cell. *See* EX2001, ¶155. Instead, he contends that “[f]or the same reasons I discussed above . . . for *Parameswaran* and *Binladen*, a POSA would not have been motivated to combine *Parameswaran* with *Binladen* and *Bentley*.” *Id.* Dr. Adalsteinsson does not dispute that, if there were a motivation to combine these references, a POSA would have had a reasonable expectation of success in doing so. *See id.*

139. As explained above, a POSA would have been motivated to combine *Parameswaran* with *Binladen*. *See supra* § VIII.A. A POSA would also have been motivated to combine *Bentley* with *Parameswaran* and *Binladen*, because a POSA would have been motivated to use the dual-indexing methods of *Parameswaran* and *Binladen* on an Illumina sequencer to reduce error rates by identifying and excluding sequences with only one identifier or incorrect pairs of identifiers. EX1003, ¶254. In December 2010, a POSA would have had a reasonable expectation of success in doing so because adopting the dual-indexing methods in *Parameswaran* and *Binladen* to an Illumina sequencer entails routine, minor modifications (e.g., swapping sequencer adaptors) without altering the fundamental aspect of the

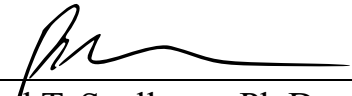
methods. *Id.*, ¶255. Indeed, switching between sequencers for various applications was routine. As such, a POSA would have found that claim 6 is obvious over *Parameswaran*, *Binladen*, and *Bentley*.

X. CONCLUSION

140. For the reasons discussed in my previous declaration and above, it is my opinion that claims 1-6 of the '200 patent are anticipated and would have been obvious to a POSA as of December 23, 2010.

141. 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. 11,768,200.
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).
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
1029	J. Binladen, et al., <i>The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing</i> , 2 PLOS ONE 2:e197 (2007).
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).
1037	International Patent Publication No. WO 2010/056728.
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 Somatically 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).

Exhibit	Description
1065	K.H. Taylor et al., <i>Next Generation Sequencing: Advances in Characterizing the Methyloome</i> , 1 GENES 143-165 (2010).
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).

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