

Filed on behalf of: Pacific Biosciences of California, Inc.

By: Michael T. Rosato (mrosato@wsgr.com)
Sonja R. Gerrard (sgerrard@wsgr.com)
Patrick M. Medley (pmedley@wsgr.com)

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

PACIFIC BIOSCIENCES OF CALIFORNIA, INC.,
Petitioner,

v.

THE CHINESE UNIVERSITY OF HONG
KONG,
Patent Owner.

Case No. IPR2024-00028
Patent No. 11,091,794

**PETITION FOR INTER PARTES REVIEW OF
U.S. PATENT NO. 11,091,794**

TABLE OF CONTENTS

	<u>Page</u>
I. Introduction.....	1
A. Background	3
1. Third-Generation Sequencing.....	4
2. SMRT Sequencing and Metrics Predictive of Modified Bases	5
3. Predictive Models	8
4. Deep Learning and Computing Power.....	10
B. Cited Art.....	11
1. The Flusberg Publications.....	11
2. PacBio-BAM.....	12
3. Clement	13
4. PacBio Template Guide	13
5. Tsai.....	14
C. The '794 Patent	14
D. Prosecution History	18
E. Denial Under §325(d) is Not Appropriate Here.....	22
1. Prong One	22
2. Prong Two.....	24
F. Level of Skill in the Art.....	26
II. Statement of the Precise Relief Requested for Each Claim Challenged	26
III. Claim Construction.....	27
IV. Detailed Explanation for Grounds of Unpatentability	27
A. [Ground 1] Claims 1-2, 7, 10-12, 14-15, and 17-19 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications in view of PacBio-BAM.....	27
1. Rationale to Reach the Claims with Reasonable Expectation of Success	29

2.	Claim 1	33
3.	Claim 2	51
4.	Claim 7	52
5.	Claim 10	53
6.	Claim 11	54
7.	Claim 12	55
8.	Claims 14 and 17	55
9.	Claim 15	57
10.	Claim 18	57
11.	Claim 19	58
B.	[Ground 2] Claims 3-6 and 9 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications, PacBio-BAM, and Clement	58
1.	Claims 3-4	60
2.	Claim 5	61
3.	Claim 6	62
4.	Claim 9	63
C.	[Ground 3] Claims 8 and 15-16 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications, PacBio- BAM, and the PacBio Template Guide	65
1.	Claim 8	66
2.	Claims 15-16	67
D.	[Ground 4] Claims 13 and 15-16 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications, PacBio- BAM, and Tsai	68
1.	Claim 13	69
2.	Claims 15-16	71
V.	There Are No <i>Fintiv</i> Issues	72
VI.	Grounds for Standing	72
VII.	Conclusion	73

VIII.	Mandatory Notices under 37 C.F.R. §42.8.....	74
IX.	Certificate of Compliance.....	74
X.	Payment of Fees under 37 C.F.R. §§42.15(a) and 42.103	76
XI.	Appendix – List of Exhibits.....	77

LISTING OF CHALLENGED CLAIMS

1. [**Preamble**] A method for detecting a modification of a nucleotide in a nucleic acid molecule, the method comprising:

[**Element 1.1**] (a) receiving data acquired by measuring pulses in an optical signal corresponding to nucleotides sequenced in a sample nucleic acid molecule and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

[**Element 1.2**] (b) creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

[**Element 1.3**] (c) inputting the input data structure into a model, the model trained by:

[**Element 1.4**] receiving a first plurality of first data structures, each first data structure of the first plurality of data structures corresponding to a respective window of nucleotides sequenced in a respective nucleic acid

molecule of a plurality of first nucleic acid molecules, wherein each of the first nucleic acid molecules is sequenced by measuring pulses in the optical signal corresponding to the nucleotides, wherein the modification has a known first state in a nucleotide at a target position in each window of each first nucleic acid molecule, each first data structure comprising values for the same properties as the input data structure,

[**Element 1.5**] storing a plurality of first training samples, each including one of the first plurality of first data structures and a first label indicating the first state of the nucleotide at the target position, and

[**Element 1.6**] optimizing, using the plurality of first training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels of the first labels when the first plurality of first data structures is input to the model, wherein an output of the model specifies whether the nucleotide at the target position in the respective window has the modification,

[**Element 1.7**] (d) determining, using the model, whether the modification is present in a nucleotide at the target position within the window in the input data structure.

2. The method of claim 1, wherein:

the input data structure is one input data structure of a plurality of input data structures,

the sample nucleic acid molecule is one sample nucleic acid molecule of a plurality of sample nucleic acid molecules,

the plurality of sample nucleic acid molecules are obtained from a biological sample of a subject, and

each input data structure corresponds to a respective window of nucleotides sequenced in a respective sample nucleic acid molecule of the plurality of sample nucleic acid molecules, and

the method further comprising:

receiving the plurality of input data structures,

inputting the plurality of input data structures into the model, and

determining, using the model, whether a modification is present in a nucleotide at a target location in the respective window of each input data structure.

3. The method of claim 2, further comprising:

determining the modification is present at one or more nucleotides,
and

determining a classification of a disorder using the presence of the modification at one or more nucleotides.

4. The method of claim 3, wherein the disorder comprises cancer.

5. The method of claim 4, further comprising:

determining that the classification of the disorder is that the subject has the disorder, and

treating the subject for the disorder by chemotherapy, radiation, or surgery.

6. The method of claim 3, wherein determining the classification of the disorder uses the number of modifications or the sites of the modifications.

7. The method of claim 2, wherein the modification is a methylation,

the method further comprising:

determining the modification is present at one or more nucleotides,
and

determining a clinically-relevant DNA fraction, a fetal methylation profile, a maternal methylation profile, a presence of an imprinting gene region, or a tissue of origin using the presence of the modification at one or more nucleotides.

8. The method of claim 2, wherein each sample nucleic acid molecule of the plurality of sample nucleic acid molecules has a size greater than a cutoff size.

9. The method of claim 2, wherein:

the plurality of sample nucleic acid molecules align to a plurality of genomic regions,

for each genomic region of the plurality of genomic regions:

a number of sample nucleic acid molecules is aligned to the genomic region,

the number of sample nucleic acid molecules is greater than a cutoff number.

10. The method of claim 1, further comprising sequencing the sample nucleic acid molecule.

11. The method of claim 1, wherein the model includes a machine learning model, a principal component analysis, a convolutional neural network, or a logistic regression.

12. The method of claim 1, wherein:

the window of nucleotides corresponding to the input data structure comprises nucleotides on a first strand of the sample nucleic acid molecule and nucleotides on a second strand of the sample nucleic acid molecule, and

the input data structure further comprises for each nucleotide within the window a value of a strand property, the strand property indicating the nucleotide being present on either the first strand or the second strand.

13. The method of claim 12, wherein the sample nucleic acid molecule is a circular DNA molecule formed by:

cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and

ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule.

14. The method of claim 1, wherein the nucleotides within the window are determined using a circular consensus sequence and without alignment of the sequenced nucleotides to a reference genome.

15. The method of claim 1, wherein each nucleotide within the window is enriched or filtered.

16. The method of claim 15, wherein each nucleotide within the window is enriched by:

cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule, or

filtered by:

selecting double-stranded DNA molecules having a size with a size range.

17. The method of claim 1, wherein nucleotides within the window are determined without using a circular consensus sequence and without alignment of the sequenced nucleotides to a reference genome.

18. The method of claim 1, wherein the optical signal is a fluorescence signal from a dye-labeled nucleotide.

19. The method of claim 1, wherein each window associated with the first plurality of data structures comprises 4 consecutive nucleotides on a first strand of each first nucleic acid molecule.

Pacific Biosciences of California, Inc., (“Petitioner” or “PacBio”) hereby requests *inter partes* review of United States Patent No. 11,091,794. This petition demonstrates by a preponderance of the evidence, that claims 1-19 of the ’794 patent are unpatentable under 35 U.S.C. §103.

I. INTRODUCTION

Years before the ’794 patent, PacBio disclosed single-molecule real-time (SMRT) sequencing which allows for simultaneous sequencing of a large array of individual nucleic acid (e.g., DNA) molecules. Sequence data is obtained by optically monitoring a polymerase as it incorporates fluorescently labeled nucleotides into a growing complementary chain. An optical “pulse” associated with a nucleotide begins when it binds to the active site of the polymerase and ends when the fluorophore is removed coincident with its incorporation into the growing chain. This process allows for sequential observation of nucleotide incorporation as the polymerase moves along the template molecule. EX1002, ¶21.

Since the early days of SMRT sequencing, the prior art literature recognized that the optical pulses provided more than just the sequence of the template molecule. Specifically, polymerase kinetics were observed from the pattern and timing of sequential pulses. Polymerase kinetic measures included the time the fluorescent nucleotide spent at the active site—termed pulse width (PW)—and the time between sequential nucleotide binding—termed interpulse duration (IPD). It

was appreciated and well-documented in the literature that the structure of the template molecule, including its sequence and any epigenetic modifications (e.g., methylation), reproducibly affected PW and IPD. EX1002, ¶22.

Long before the '794 patent, sequence data and polymerase kinetic measures obtained in the course of SMRT sequencing were used to predict sites of epigenetic modification. Indeed, this was one of the well-recognized advantages of SMRT sequencing relative to older sequencing technologies (e.g., Illumina sequencers). In this regard, skilled artisans developed models (e.g., machine learning) to analyze SMRT pulse data and characterized the effects of sequence and epigenetic modification on PW and IPD. *Id.*

The '794 patent broadly claims use of these long-recognized parameters for detecting base modifications in SMRT pulse data using a “model.” The '794 patent does not assert any novelty in the claimed SMRT pulse data parameters nor could it. The prior art discloses each of these parameters—sequence context (identity and position), PW, and IPD—are conventionally generated in the course of SMRT sequencing and used to predict base modification. The '794 patent does not assert any novelty in the generically claimed “model.” Indeed, the specification acknowledges the claims broadly encompass statistical, machine learning, and other types of prior art models used to analyze genomic data including that obtained on the SMRT platform. EX1002, ¶23.

PO gained allowance of the challenged claims by falsely representing that its “data structure” (nucleotide identity, position, PW, IPD) was a point of novelty. However, art withheld from the Office demonstrates the falsity of its arguments. Indeed, PO’s own briefing from district court, in arguing mapping to the claims, identifies a data structure identically found in the published prior art long before the ’794 patent. As this petition demonstrates, corroborated by PO’s own argument, the input data structure of the claims is provided by the BAM file which the prior art describes as generated in the course of SMRT sequencing. EX1002, ¶24.

Claims 1-19 of the ’794 patent should be found unpatentable and cancelled.

A. Background

This petition is supported by the Declaration of Dr. Christophe Dessimoz the Executive Director of the Swiss Institute of Bioinformatics and Associate Professor in the Department of Computational Biology at the University of Lausanne. EX1002, ¶¶1-11; EX1003. As Dr. Dessimoz explains and as addressed in further detail below, the challenged claims would not have been considered new or non-obvious to a person of ordinary skill in the art (“POSA”) at the relevant time.

1. Third-Generation Sequencing

Third-generation sequencing (e.g., SMRT) was known and commercially available ~9 years prior to the '794 patent's earliest claimed priority date. EX1026, 1. In contrast to older technologies, in SMRT sequencing, a single DNA polymerase enzyme is observed in real time as it adds nucleotides to a growing strand.¹ Because the enzyme processing rate was being recorded, it was recognized that epigenetic base modifications (e.g., methylation) may be detected along with sequence data. EX1025, 2344-46. Epigenetic modification plays a role in in many critical biologic processes including gene expression, development, and cancer, *e.g.*, EX1007, 1:39-2:10, 15:26-16:59, thus there has been considerable interest in using third-generation technologies for modification detection. *E.g.*, EX1006, 461-462; EX1007, Abstract; EX1008, Abstract; EX1009, Abstract; EX1010, Abstract; EX1015, Abstract; EX1016, 1; EX1022, Abstract; EX1025, 2345; EX1027, 18910; EX1035, Abstract; EX1038, Abstract; *see* EX1039, EX1041. To this end, modification detection methodologies have been developed for PacBio's SMRT

¹ A video briefly summarizing SMRT sequencing, corresponding to the disclosure of EX1023, may be viewed at <https://www.pacb.com/technology/hifi-sequencing/how-it-works/> (EX1044). EX1002, ¶25.

and Oxford Nanopore Technologies (“ONT”) platforms. EX1002, ¶¶25-26; *see* §I.A.2.

2. SMRT Sequencing and Metrics Predictive of Modified Bases

Prior to the ’794 patent, SMRT sequencing obtained sequence and polymerase kinetic data by optically monitoring a single polymerase as it incorporates nucleotides into a growing complementary strand. As shown below, SMRT sequencing generates a series of fluorescent pulses in which the color of each pulse indicates the identity of the incorporated nucleotide. The sequential pulses also provide kinetic parameters, including the width of the pulse (PW) and the time between pulses (IPD). EX1006, Fig. 1; *see* EX1007, Fig. 2; EX1009, 127, Fig. 1; EX1013, 194; EX1002, ¶¶27-28.

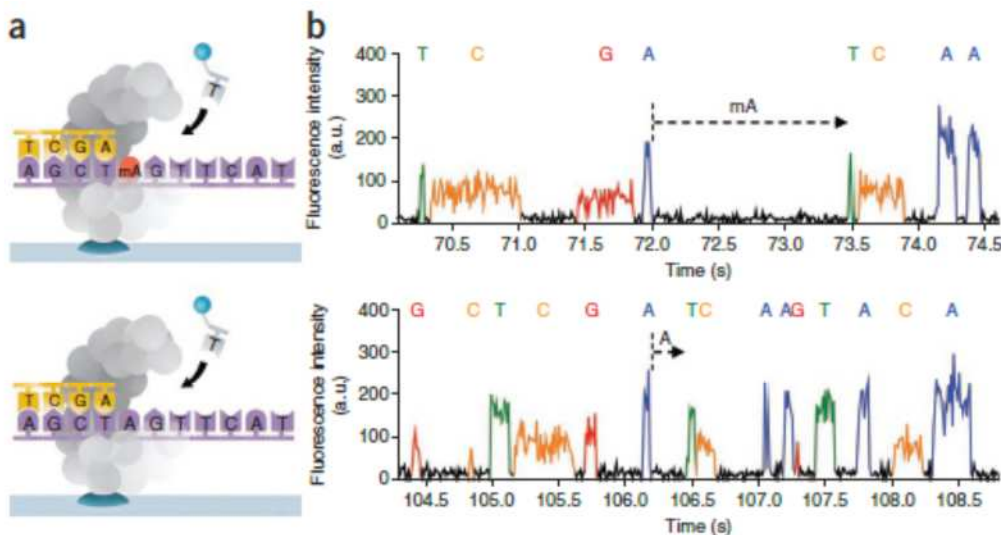


Figure 1 | Principle and corresponding example of detecting DNA methylation during SMRT sequencing. (a) Schematics of polymerase synthesis of DNA strands containing a methylated (top) or unmethylated (bottom) adenine. (b) Typical SMRT sequencing fluorescence traces for samples in a. Letters above the fluorescence trace pulses indicate the identity of the nucleotide incorporated into the growing complementary strand. Dashed arrows indicate the IPD before incorporation of the cognate thymine. For this typical example, the IPD is about five times larger for mA in the template compared to adenine.

Polymerase kinetic parameters, such as PW and IPD, were known to provide valuable information regarding base modifications.

Fluorescence pulses in SMRT sequencing are characterized not only by their emission spectra but also by their duration and by the interval between successive pulses. These metrics, defined here as pulse width and interpulse duration (IPD), add valuable information about DNA polymerase kinetics. Pulse width is a function of all kinetic steps after nucleotide binding and up to fluorophore release, and IPD is determined by the kinetics of nucleotide binding and polymerase translocation.

EX1006, 462 (modifications affect kinetics); *see* EX1007, 22:25-54; EX1002, ¶29.

It was further known and well-documented in the literature that polymerase kinetics were strongly affected by sequence context including the bases upstream and downstream of an incorporated base.

Another interesting feature of the DNA polymerase kinetics in SMRT sequencing is that it is exquisitely sensitive to local sequence context.... In fact, seven bases upstream of and two bases downstream from the incorporation site explain >80% of the variance in the IPD measures (Fig. 2A).

EX1010, 130; *see* EX1006, 462-463, 476; EX1007, 62:29-40; EX1008, 2; EX1009, 124; EX1002, ¶30.

POSAs also appreciated along with sequence information, base modifications affected polymerase kinetics (such as reflected in PW and IPD), which could be used to predict modification. *E.g.*, EX1006, 463 (“[E]ach modification had a unique IPD and pulse width signature in a given context...”), Fig. 3; EX1007, 22:25-54 (“[T]he presence of modified bases in a template nucleic acid has been shown to change the width of a pulse and/or the interpulse duration (IPD)...”); EX1002, ¶¶31-32. *See generally* EX1008. For example, the presence of a modified base was known to alter polymerase kinetics at nearby bases in a manner characteristic of the type of base modification, further illustrating the known significance relationship between sequence context and kinetic parameters.

Kinetic effects are not necessarily limited to just the nucleotide incorporation opposite the modified base position in the DNA template. This is because the DNA polymerase is in intimate contact with the DNA over an extended region of approximately twelve bases, and the modified base can impart effects on the polymerase dynamics at several positions over this region. This results in kinetic “signatures” which can aid in identifying the type of base modification.

EX1016, 3; *see* EX1010, 130; EX1006, 462-464, Fig. 2, 476; EX1007, 4:66-5:3, 22:25-54, 78:18-43, 79:2-5, Figs. 18-19; EX1008, 2; EX1002, ¶¶33.

Prior to the '794 patent, pulse data generated from SMRT sequencing were stored together in a data structure referred to as a “BAM file.” EX1011, 1.

PacBio’s BAM file documentation published prior to the '794 patent teaches that a PacBio BAM file conventionally stored sequence data and kinetic parameters (e.g., PW and IPD) for each nucleotide of the template. EX1011, 1-2, 4-5, 8. This documentation also describes PacBio’s BAM files as “fully compatible with the BAM specification” and specifies that a BAM file includes the identity of each nucleotide, in sequence, as an array. EX1011, 1; *see* EX1012, 15; EX1002, ¶¶34-35.

3. Predictive Models

Prior to the '794 patent, metrics predictive of modification state were known, as was inputting these metrics into models to train them to identify

modified bases. The '794 patent acknowledges as much in its discussion of prior art models such as deep learning, hierarchical, conditional random field, and principal component analysis to identify modified bases in data generated from SMRT and ONT platforms. EX1001, 17:32-35, 19:2-11, 19:14-18, 35:21-37; *see Qualcomm Inc. v. Apple Inc.*, 24 F.4th 1367, 1376 (Fed. Cir. 2022) (“[P]atentee’s admissions about the scope and content of the prior art provide a factual foundation as to what a skilled artisan would have known at the time of invention...”); EX1002, ¶36.

The prior art taught multiple different models that used third-generation sequencing data to detect base modifications. For example, PacBio’s kineticsTools manual describes a model that receives a BAM file as an input and uses the information in that file, including IPD values and a window of nucleotides to predict whether particular bases are modified. EX1021, 1-2; EX1002, ¶37. Other references describe models that use a combination of kinetic parameters (including PW and IPD) and the local sequence context over a window of nucleotides. EX1010, 130; EX1009, 124, 126; EX1002, ¶38.

Other third-generation sequencing technologies (e.g., ONT sequencers) also used a combination of kinetic metrics and sequence context for predicting base modification. For example, Liu teaches using a trained neural network to determine methylation from ONT sequencing data. EX1022, Abstract, 3. Q. Liu describes

using a data structure containing kinetic features and nucleotide identities for each nucleotide in a window centered around a nucleotide of interest, with the output indicating whether the base in the center of the window is modified. EX1022, 8-9, Fig. 1; *see* EX1038, 2 (With PacBio and ONT data, “modifications [are detected] by measuring the kinetics.”); EX1002, ¶¶39-41. Other models, such as support vector machines, were also used for prediction of methylation in ONT sequencing data. EX1038, 3. As Dr. Dessimoz explains, a POSA would have known that, regardless of sequencing platform, prediction of methylation using models would involve training with features associated with methylation. EX1002, ¶41.

4. Deep Learning and Computing Power

The commercial viability of machine learning techniques was enhanced in the decade following the publication of Flusberg (EX1006) due at least in part to advancements in computing power and deep learning networks (e.g., neural networks). As explained in a 2019 report, “computing power has increased by orders of magnitude” in the past decade. EX1028, 787; *see id.* (discussing cloud-based services). These improvements coincided with rapid development of deep learning networks which required high-performance computing to train and run. EX1031, 2; EX1002, ¶42.

While machine learning was known for decades, more widespread implementation of such technologies in genomic applications paralleled the

increased availability of open-source resources. EX1028, 782; EX1017, 16-17; EX1035, 2; EX1032; EX1036, 1; EX1038, 3-4; EX1030, 10; EX1002, ¶¶43-45.

B. Cited Art

1. The Flusberg Publications

Flusberg et al. published “Direct detection of DNA methylation during single-molecule, real-time sequencing” (“Flusberg,” EX1006) in the online version of *Nature Methods* on May 9, 2010. Flusberg is prior art under 35 U.S.C. §102(a)(1). U.S. Patent No. 9,175,338 to Flusberg, et al. is entitled “Methods for identifying nucleic acid modifications” (“Flusberg-338,” EX1007). Flusberg-338 was filed on November 12, 2010, issued on November 3, 2015, and incorporates Flusberg by reference. EX1007, 77:30-34, 78:58-66. Flusberg-338 is prior art under 35 U.S.C. §§102(a)(1) and (a)(2). Flusberg and Flusberg-338 are collectively referred to herein as “the Flusberg Publications.”

The Flusberg Publications disclose that sequence context and base modifications alter polymerase kinetics in a predictable fashion. These publications further disclose that polymerase kinetics were reflected in the PW and IPD metrics which are obtained along with base identity from SMRT sequencing. *E.g.*, EX1006, 462-463, 476; EX1007, 22:25-54, 62:29-38; *supra* §I.A.2.

The Flusberg Publications further describe multiple models (e.g., statistical, machine learning) for identifying modified bases (e.g., methylation) in template

molecules using data obtained from SMRT sequencing. *E.g.*, EX1006, 462-463, Fig. 3; EX1007, 61:6-69:51. For example, the Flusberg Publications disclose using models to identify modified bases by comparing PW and IPD values obtained for a potentially modified nucleic acid molecule to values obtained for an unmodified control of the same sequence. EX1006, Figs. 1, 3, 463-65, 476; EX1007, 78:18-43. Flusberg-338 also discloses training machine learning models with data (e.g., sequence context, PW, and IPD) obtained from sample molecules having a known modification state and using the trained model to identify modified bases. EX1006, 463-65; EX1007, 64:29-48, 66:46-67:36; EX1002, ¶¶46-50.

2. PacBio-BAM

The PacBio BAM Format Documentation (PacBio-BAM, EX1011) was published by PacBio on its website on or before September 16, 2017. PacBio-BAM's public availability is confirmed by an Internet Archive WayBack Machine capture of PacBio-BAM on September 16, 2017. EX1040; *see* EX1002, ¶54 (also publicly available via GitHub); EX1042; EX1043. PacBio-BAM is prior art under 35 U.S.C. §102(a)(1).

PacBio-BAM discloses that data obtained from SMRT sequencing (e.g., identity, position, PW, and IPD values) are stored in a BAM file. PacBio-BAM also discloses that the PacBio BAM file stores the identity of the bases in order. EX1011, 1-2; *see* EX1012, 5, 15; EX1040 (linked EX1012 archived Jun. 13,

2018). PacBio-BAM further discloses that values for pulse features (e.g., PW and IPD) are stored in the PacBio BAM file in a manner compatible with downstream analyses including “base modification detection.” EX1011, 4-5, 8; EX1002, ¶¶51-53.

PO confirms these disclosures in its infringement contentions. EX1014, 23-35; *supra* §I.D.

3. Clement

U.S. Patent Publ. No. 2016/0326593 to Clement et al. (“Clement,” EX1018) is entitled “Compositions and Methods for Diagnosing, Evaluating and Treating Cancer.” Clement was filed on May 23, 2016 and published on November 10, 2016. Clement is prior art under 35 U.S.C. §§102(a)(1) and (a)(2).

Clement discloses “methods of determining a cancer treatment prognosis” based on “evaluating epigenetic and genetic changes within a tumor sample.” EX1018, Abstract. Clement describes classifying samples according to methylation pattern. EX1018, ¶¶9, 78, 198; EX1002, ¶¶55-56.

4. PacBio Template Guide

The PacBio Template Guide (EX1019) was published by PacBio in January 2015 and was distributed to its customers prior to the ’794 patent filing. EX1041. The PacBio Template Guide’s public availability is confirmed by an Internet

Archive WayBack Machine capture of the document on February 7, 2016.

EX1039. The PacBio Template Guide is prior art under 35 U.S.C. §102(a)(1).

The PacBio Template Guide discloses preparation of DNA template libraries for sequencing on the SMRT platform. It further discloses selecting molecules for sequencing within a preferred size range and removal of molecules below a defined size. EX1019, 24-28, 35; EX1002, ¶¶57-58.

5. Tsai

U.S. Patent Publication No. 2016/0208241 to Tsai et al. (“Tsai,” EX1020) is entitled “Compositions and Methods for Enrichment of Nucleic Acids.” Tsai was filed on January 14, 2016 and published on July 21, 2016. Tsai is prior art under 35 U.S.C. §102(a)(1) and (a)(2).

Tsai describes an amplification-free enrichment technique using the CRISPR-Cas9 system to generate SMRT sequencing libraries (EX1020, ¶¶6-7, 65-66, Fig. 8) using PacBio SMRT sequencing, teaching that “[p]referably the sequencing method is capable of detecting base modifications during a sequencing reaction, e.g., by monitoring the kinetics of the sequencing reaction.” EX1020, ¶23. EX1002, ¶¶59-60.

C. The '794 Patent

The '794 patent is entitled “Determination of Base Modifications of Nucleic Acids.” EX1001, cover. The '794 patent simply follows teaching that has long

been known and documented in the prior art—using sequence context and kinetic metrics (e.g., PW and IPD) obtained from PacBio’s SMRT platform to detect base modifications such as methylcytosine. EX1001, 2:54-64, 19:54-60; EX1002, ¶61.

Claim 1 requires acquiring data generated in the course of SMRT sequencing, storing it in an input data structure (e.g., BAM file), and then inputting the stored data into a trained model to determine modified bases. The recited properties that are acquired, stored, and input into a model include sequence context (“identity” and “position”) and metrics associated with polymerase kinetics (“width of the pulse” and “interpulse duration”). EX1002, ¶62.

Virtually every aspect of the claimed subject matter is expressly identified in the ’794 patent as being known in the art. The ’794 patent acknowledges that third-generation sequencing platforms have long been used to detect modified bases. EX1001, 13:7-17 (citing Flusberg (EX1006) and Schreiber (EX1027)). The ’794 patent also acknowledges that sequence context and polymerase kinetics (e.g., IPD and PW) were previously used to predict base modification. EX1001, 19:2-13, 19:38-40. *See generally* EX1001, 17:7-19:50; EX1002, ¶63.

The ’794 patent, in fact, credits Flusberg for demonstrating “polymerase kinetics would be affected by methylation states in the sequences.” EX1001, 17:64-65. The ’794 patent acknowledges that Flusberg analyzed polymerase kinetics, as reflected in PW and IPD, to detect various base modifications.

EX1001, 18:19-30; *see* EX1006, 462 (“These metrics, defined here as pulse width and interpulse duration (IPD), add valuable information about DNA polymerase kinetics.”).²

The ’794 patent states that a broad array of art-known models are within the scope of the invention. EX1001, 26:34-45 (“[A]ny model or data analysis framework....”), 72:53-55, 89:5-51. For example, the ’794 patent states that models within the scope of the invention include conditional random field, deep learning, hierarchical models, and principal component analysis. EX1001, 29:39-43, 77:39-67, 89:25-51, claim 11. The ’794 patent acknowledges that these same models had long been used for detecting modified bases in third-generation sequencing data. EX1001, 17:32-35, 19:2-11, 19:14-18, 35:21-37. However, just as the ’794 patent discloses use of input data structures for training models (*see* EX1001, 87:42-88:3, 90:20-91:31, Figs. 102, 103), POSAs would have used data structures in the prior art implementations of those same models. EX1002, ¶66.

² The ’794 patent asserts Flusberg suggests that PW is not significantly affected by base modifications. EX1001, 18:19-30. But Flusberg expressly states that base modifications affect PW. *E.g.*, EX1006, 463 (“[E]ach modification had a unique IPD and pulse width signature in a given context.”), Fig. 3; EX1002, ¶65.

To the extent the '794 patent alleges a point of novelty, it is the combined use of multiple parameters (sequence context, PW, and IPD) that were admittedly and indisputably known for detection of base modifications.

We reasoned that if we can use a combination of multiple metrics, for example, concurrently making use of features including upstream and downstream sequence context, strand information, IPD, pulse widths as well as pulse strength, we might be able to achieve the accurate measurement of base modifications (e.g. mC detection) at single-base resolution.

EX1001, 19:54-60. However, it is well-documented in the prior art—and acknowledged in the '794 patent—that each of these metrics was predictive of modification state. Furthermore, as established by prior art not before the Office, values for each of these metrics were routinely generated in the ordinary course of SMRT sequencing, included in PacBio's generated BAM file, and users were instructed to use kinetic data in the BAM file for modification detection. As demonstrated herein, making use of multiple metrics known to be predictive of modification state which were routinely obtained from SMRT sequencing was known or, at a minimum, would have been obvious to a POSA in view of the prior art. *Supra* §§I.A, IV; EX1002, ¶¶67-68.

D. Prosecution History

The '794 patent was filed as U.S. Application No. 16/995,607 (“the ’607 application”) on August 17, 2020, claiming priority to several Provisional Patent Applications filed as early as August 16, 2019.

In the only Office action, original claims 1-17, 70, and 71 were rejected under §101 as directed to “mathematically manipulating and comparing previously obtained optical pulse sequencing data.” EX1004, 800-02, 1090-94 (original claims). The claims were also rejected as anticipated by Flusberg and obvious over Flusberg-338. EX1004, 803-10.

PO initiated an interview discussing amending claim 1 to include “creating an input data structure....” EX1004, 217. The Examiner also summarized that she did not understand the prior art to describe an input data structure containing the listed properties. EX1004, 217.

PO responded by amending claim 1 to recite “creating an input data structure...” and falsely arguing it represents a point of novelty.

The input data structure is not well-understood, routine, or conventional at least because obtaining all the values for the listed properties is not well-understood, routine, or conventional, let alone combining into an input data structure and inputting into a model. As a result, “receiving data acquired by measuring pulses in an optical signal,” “obtaining, from the data, values for” certain properties, “creating an input data

structure” and “inputting the input data structure into a model” are not well-understood, routine, or conventional activity alone or in combination with other limitations of the claim.

EX1004, 206. PO argued that neither Flusberg Publication describes an input data structure containing all listed properties. EX1004, 206-209. A notice of allowance followed but offered no reasons for allowance. EX1004, 174-178.

During prosecution of a continuation application, the same Examiner found the closest prior art references were PacBio-owned patents including Flusberg-338. EX1005, 150. While issuing no prior art rejections, the Examiner’s notice of allowance accepted PO’s prior arguments regarding the novelty of the input data structure:

None of the US Patents demonstrate measuring a pulse width value of a specifically identified nucleotide at an identified position to determine the methylation status, and do not teach or suggest combining all data measurements of nucleotide identities, nucleotide positions in the sample nucleic acid, width of the pulse corresponding to the nucleotides, and IPD, subsequently creating an input data structure comprising a window of the nucleotides in a sample that lists the above measurements, and inputting the measurements into a model configured to utilize the measurements to output a methylation status for each nucleotide as instantly claimed.

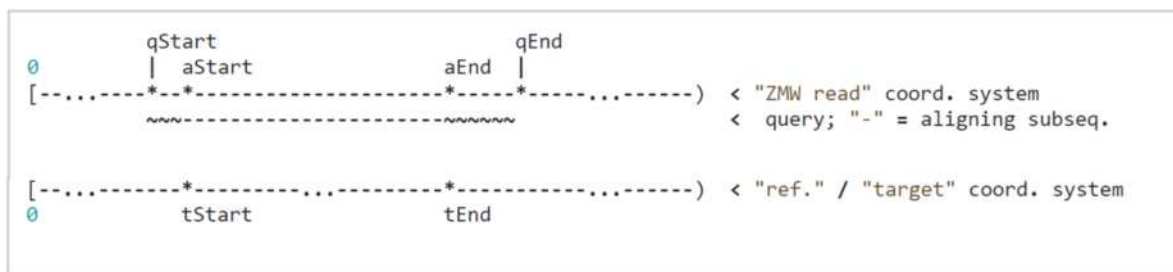
EX1005, 150.

The claimed “input data structure,” however, is not novel. PacBio’s prior art BAM file documentation (PacBio-BAM, EX1011) establishes that the BAM file generated in the course of SMRT sequencing provides a data structure containing the required properties. PO can dispute none of this because its infringement contentions (EX1014) rely on substantially identical disclosure found in a later version of the BAM file documentation (*see* EX1033, 5-16) to allege infringement of the same claims.

Specifically, with respect to “creating an input data structure,” PO’s infringement contentions allege PacBio products store data generated by PacBio’s SMRT sequencing platform in a BAM file. *E.g.*, EX1014, 23, 25.

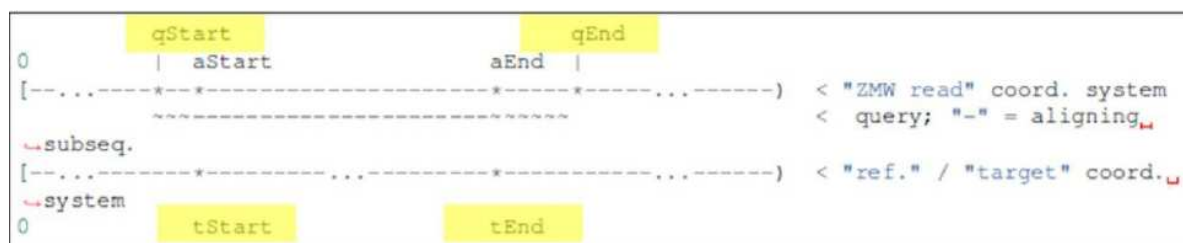
As confirmed by Dr. Dessimoz, the same content that PO relies on in the 2022 version of the BAM file documentation (PacBioFileFormats-2022, EX1033) to allege infringement is also found in a prior art version of the documentation (PacBio-BAM). EX1002, ¶¶69-73. For example, PO alleges the “identity of the nucleotide” is found “in the SEQ field” of PacBioFileFormats-2022. EX1014, 28. The SEQ field is similarly provided in the prior art PacBio-BAM. EX1011, 8 (discussing the SEQ field); *see* EX1012, 5 (SEQ is a “mandatory field[]”). PO alleges, relying on the figure below, that the “position of the nucleotide” is provided in PacBioFileFormats-2022’s BAM file. This disclosure is found identically in the prior art PacBio-BAM.

Prior art BAM file:



In our BAM files, the qStart, qEnd are contained in the q_s and q_e tags, (and reflected in the q_{NAME}); the bounds of the *aligned query* in the ZMW read can be determined by adjusting q_s and q_e by the number of soft-clipped bases at the ends of the alignment (as found in the CIGAR).

Allegedly infringing BAM file:



In our BAM files, the qStart, qEnd are contained in the qs and qe tags, (and reflected in the QNAME); the bounds of the aligned query in the ZMW read can be determined by adjusting qs and qe by the number of soft-clipped bases at the ends of the alignment (as found in the CIGAR).

EX1014, 30; EX1011, 1-2; *see* EX1033, 6. PO alleges that values for PW and IPD are contained in PacBioFileFormats-2022's BAM file. EX1014, 32, 34; *see* EX1033, 11. The prior art PacBio-BAM also discloses storing values for PW and IPD. EX1011, 4-5, 8.

Finally, PO alleges PacBioFileFormats-2022's BAM file is used as input into PacBio's accused base modification caller.

Required	Description
HiFi INPUT	Input BAM file or ConsensusReadSet XML file.
HiFi OUTPUT	Output BAM file or ConsensusReadSet XML file.

EX1014, 24. The prior art PacBio-BAM expressly discusses using the BAM file for “base modification detection.” EX1011, 8. Further, regarding the data structure being an *input* data structure for a base modification model, PacBio’s prior art base modification caller (kineticsTools) used a BAM file as input. EX1021, 1-2.

Accordingly, the claimed “input data structure” is not novel. Had the Examiner been aware of PO’s district court infringement contentions or that PacBio prior art disclosed using a BAM file containing the listed properties long before the ’794 patent, she would not have had good basis to allow the case.

E. Denial Under §325(d) is Not Appropriate Here

Discretionary denial is not warranted under the Board’s two-prong §325(d) analysis. *Advanced Bionics, LLC v. Med-El Elektromedizinische Geräte GMBH*, IPR2019-01469, Paper 6 at 9-10 (precedential). First, the art and arguments presented here are distinct from what was previously considered by the Office and directly address PO’s material misrepresentations that led to the issuance of the ’794 patent. To the extent the second prong of the analysis is reached, issuance of the ’794 patent over the Flusberg Publications constitutes material error by the Office.

1. Prong One

Each of Petitioner’s four grounds rely upon art that was not before the Examiner including PacBio-BAM, Clement, Template Guide, and Tsai. *See* §IV.

The Examiner also did not consider PO's infringement contentions or Dr. Dessimoz's testimony, which further undermine PO's argument that the claimed "input data structure" was novel. Accordingly, the Office has not previously considered obviousness of the claims of the '794 patent in view of the art asserted here. *Thorne Research, Inc. v. Trs. of Dartmouth Coll.*, IPR2021-00491, Paper 18 at 7-9 ("The fact that one piece of art from the combination was previously presented and/or argued to the Office alone is insufficient to satisfy the first prong of *Advanced Bionics* two-part test....").

The new art and evidence are not cumulative and, as explained above, directly address PO's misrepresentations. *See* §I.D. PO argued that the claimed "input data structure" is novel but PacBio-BAM indisputably describes such an input data structure. EX1011, 1-9; *see Geotab, Inc. v. Omega Patents, LLC*, IPR2023-00504, Paper 11, 14-17 (prong one not satisfied where new art "teaches or suggests claim limitations missing from the prior art previously considered by the Office"); *Kea Cloud, Inc. v. Valyant AI, Inc.*, IPR2023-00450, Paper 12, 14-15 (prong one not satisfied where new art addresses "limitation [that] was added during prosecution and formed a key basis for Patent Owner's arguments over the art of record").

Additional references not before the Examiner further confirm obviousness. For example, Clark, Loomis, and the PacBio White Paper reinforce the teachings

of the Flusberg Publications that sequence context, PW and IPD were recognized variables considered when identifying base modifications. EX1008, 2; EX1009, 4 (citing EX1006); EX1016, 3. The kineticsTools manual establishes that a BAM file was used as input into a model for detection of base modifications. EX1021, 1-2; *see* EX1011, 8 (BAM file suitable for “base modification detection”).

These additional references confirm that a POSA would have been motivated and would have reasonably expected success in using base identity, position, PW, and IPD from PacBio’s BAM file to identify base modifications as claimed. *Crowdstrike, Inc. v. Open Text Inc.*, IPR2023-00289, Paper 7, 20-21 (prong one not satisfied where it was “not apparent that the Examiner considered whether it would have been obvious to modify [the prior art] to incorporate [a feature], particularly in light of the teachings of an additional reference that was not before the Examiner”).

2. Prong Two

To the extent prong two is reached, the record demonstrates the Office committed legal and factual error in issuing the ’794 patent.

First, as discussed above, allowance of the ’794 patent was material error where PO argued that the claimed “input data structure” is novel.

Second, the Office committed legal error by directly applying the reasoning from the Flusberg anticipation rejection to the Flusberg-338 obviousness rejection.

EX1004, 217 (“35 U.S.C. 103 See comments in 102 above.”). In so doing, the Office appears to have evaluated Flusberg-338 under an anticipation theory rather than considering obviousness. As established in Grounds 1-4 of this petition, the claims of the ’794 patent were obvious in view of combinations including Flusberg-338. *Supra* §IV.

Third, the Examiner committed factual error to the extent she found that the prior art does not teach “measuring a pulse width value of a specifically identified nucleotide at an identified position to determine the methylation status.” EX1005, 150. Flusberg-338 has express teachings on point. *E.g.*, EX1007, 12:63-65 (“FIG. 19 provides a graph plotting pulse width ratio against template position for a template nucleic acid comprising 8-oxoguanosine modifications.”), 35:13-16 (“...monitoring deviations from normal polymerase kinetics, including IPD and pulse width.”), 68:22-37 (“parameters of interest (e.g., pulse width, IPD, etc.)”), 78:36-43 (“8-oxoG altered pulse width over a window of 7-8 neighboring bases”), Fig. 19. Moreover, Flusberg, which is incorporated by reference into Flusberg-338, teaches measuring PW at identified positions to determine both modification status and the identity of the modification. *E.g.*, EX1006, Fig. 3 (“Each principal component is a linear combination of the mean IPD and pulse width...”). Accordingly, Flusberg-338 teaches measuring a PW value of a specifically identified nucleotide at an identified position to determine the methylation status

and it was material error to find otherwise. *See Samsung Elecs. Co. v. G+ Commc'ns, LLC*, IPR2022-01598, Paper 10, 16 (finding material error where the Examiner overlooked disclosure of the alleged missing limitation in the applied art).

F. Level of Skill in the Art

As Dr. Dessimoz explains, a POSA by August 16, 2019 would typically have a M.S. or Ph.D. degree with 2-4 years of research experience in genomics, bioinformatics, computational biology, or a related field, or could have less education but significant professional experience in one or more of these fields. Such a POSA would have possessed knowledge of the references discussed above, among other relevant prior art. EX1002, ¶¶74-75.

II. STATEMENT OF THE PRECISE RELIEF REQUESTED FOR EACH CLAIM CHALLENGED

Petitioner requests review of claims 1-19 of the '794 patent under 35 U.S.C. §311 and AIA §6. The grounds for relief are as follows:

Ground	Claims	Description
1	1-2, 7, 10-12, 14-15, 17-19	Unpatentable under 35 U.S.C. §103: Flusberg Publications and PacBio-BAM
2	3-6, 9	Unpatentable under 35 U.S.C. §103: Flusberg Publications, PacBio-BAM, and Clement
3	8, 15-16	Unpatentable under 35 U.S.C. §103: Flusberg

		Publications, PacBio-BAM, and Template Guide
4	13, 15-16	Unpatentable under 35 U.S.C. §103: Flusberg Publications, PacBio-BAM, and Tsai

III. CLAIM CONSTRUCTION

Claims are construed to the extent necessary to find the claims unpatentable in view of the prior art. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017). In mapping the prior art to the claims, Petitioner has applied the plain “meaning that [a] term would have to a person of ordinary skill in the art in question at the time of the invention.” *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (en banc). Claims 3-6, 8, 9, and 11, however, contain terms expressly defined in the ’794 patent. EX1001, 14:65-15:5 (“classification”), 15:6-24 (“cutoff”), 16:34-40 (“machine learning models”). For these claims, Petitioner has applied the definitions in the specification. EX1002, ¶¶76-77.

IV. DETAILED EXPLANATION FOR GROUNDS OF UNPATENTABILITY

A. [Ground 1] Claims 1-2, 7, 10-12, 14-15, and 17-19 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications and PacBio-BAM

As discussed in detail below, the Flusberg Publications and PacBio-BAM render claims 1-2, 7, 10-12, 14-15, and 17-19 obvious.

The Flusberg Publications disclose determining whether a nucleotide is modified, such as by being methylated. EX1006, Abstract; EX1007, Abstract. The

Flusberg Publications disclose that nucleic acid molecules are sequenced using PacBio SMRT sequencing, generating pulse traces in which each nucleotide produces a pulse from which characteristic features may be derived. EX1006, 462-63, Fig. 1; EX1007, 23:14-22, 64:29-43, 66:46-67:4, Figs. 1 & 2. These features include the identities of each nucleotide in the sequence, as well as parameters including PW and IPD for each nucleotide. *Id.* Indeed, the Office recently recognized Flusberg-338's specification (continuation application No. 18/189,084) as disclosing a method of identifying modified bases using a model with an input feature vector (e.g., data structure comprising a window) comprising basecalls, PWs, and IPDs. EX1037, 17-18, 36-39.

The Flusberg Publications teach that pulse features (e.g., PW and IPD) are tabulated for a window of nucleotides and used in a previously trained model to determine whether a nucleotide is methylated. *E.g.*, EX1006, 463, 465-466, Fig. 3; EX1007, 66:46-67:36. For example, Flusberg teaches using "principal component analysis" which takes an input data structure containing PW and IPD signals over a window of nucleotides and outputs a signal discriminating between different modification states (including unmodified) at a given nucleotide. EX1006, 463. Flusberg-338 similarly describes analytic models for determining methylation state that "can be implemented with many standard statistical classification algorithms"

and trained with pulse data (e.g., context, PW, and IPD) over a window of nucleotides. EX1007, 66:46-68:8.

To the extent the Flusberg Publications are not seen to expressly disclose each of nucleotide identity, position, PW, and IPD in an input data structure, this is taught by PacBio-BAM, which discloses the standard data structure in which PacBio SMRT sequencers store data. EX1011, 1-2. These data structures contain, *inter alia*, identity, position, PW, and IPD values for each nucleotide in a sequence. EX1011, 1-2, 4-5, 8; EX1012, 15; EX1002, ¶¶78-81.

1. Rationale to Reach the Claims with Reasonable Expectation of Success

A POSA would have had good reason to combine the Flusberg Publications with PacBio-BAM to achieve what is claimed with a reasonable expectation of success. The Flusberg Publications disclose detecting modifications using PacBio SMRT sequencing. PacBio-BAM teaches a BAM file containing values for all claimed properties that was generated in the course of SMRT sequencing. As established herein, BAM files generated by the SMRT sequencing platform contain data for the identity and position of bases along with PW and IPD. EX1011, 4-5, 8. PacBio-BAM further describes the file as compatible with “base modification detection.” EX1011, 8; *see* EX1021, 1-2 (using PacBio’s BAM file as an input data structure for base modification detection); EX1002, ¶82. A POSA would have thus

been motivated to use PacBio's BAM file containing the recited properties as an input data structure, as expressly suggested in PacBio-BAM. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 418-19 (2017) (explaining that an express suggestion in the literature, while not required to demonstrate obviousness, strongly supports a conclusion of obviousness); *see also supra* §§I.B.1-2.

Additionally, it was well-known and documented in the prior art that values for the claimed SMRT sequencing properties were routinely generated and known as useful in base modification detection. Specifically, POSAs appreciated that the sequence context (e.g., identity and positions of the bases in the vicinity of a modified base) strongly affects polymerase kinetics, and as a result, PW and IPD values. EX1006, 462-463; EX1007, 62:29-31; EX1008, 1-2; EX1009, 124; EX1016, 2; EX1010, 130. POSAs appreciated polymerase kinetics were dependent on the presence of base modifications and different modification types had unique PW and IPD signatures. EX1006, 462, Fig. 2; EX1007, 4:66-5:3, 22:45-54, 78:18-43, 79:2-5, Figs. 18-19; EX1008, 2; EX1010, 130; EX1016, 3. Because of the documented interdependence of local sequence context and base modifications on PW and IPD, POSAs appreciated that local sequence effects should be considered when predicting base modifications. *E.g.*, EX1006, 462-63 (“[K]inetic signatures of methylation will be sequence context-dependent.”), 476; EX1010, 130 (“seven bases upstream of and two bases downstream from the incorporation site explain

>80% of the variance in the IPD measures”); *supra* §I.A.2. Thus, as Dr. Dessimoz explains, a POSA would have found it obvious to consider the claimed properties together because doing so represents nothing beyond the predictable use of known variables according to their well-characterized role in base modification detection. EX1002, ¶¶83-84 *KSR*, 550 U.S. at 417 (“[A] court must ask whether the improvement is more than the predictable use of prior art elements according to their established functions.”).

Moreover, as Dr. Dessimoz explains, the scientific literature at the time further corroborates the obviousness of considering data for nucleotide identity, position, PW, and IPD together. EX1002, ¶85. For example, Loomis describes generating a file containing SMRT sequencing data for analysis of polymerase kinetics. EX1009, 127. This data file included the base identities in order linked with pulse parameters such as PW and IPD. EX1009, 127. Loomis describes performing a statistical analysis on this data, which showed “differences in the rates of polymerization” across different regions, and even individual bases, of a template molecule. EX1009, 124, 127; *see id.*, 124 (“The principal outcome measure...is the interpulse distance (IPD), ...The other outcome measure is the pulse width (PW), ...reflects a context-dependent local DNA conformation experienced by the polymerase....”), Fig. 3. While Loomis observed polymerase kinetics on unmodified DNA, the authors describe their analyses as “establish[ing]

a clear foundation for the detection of epigenetic modifications.” EX1009, 126; *see supra* §I.A.2.

Additionally, models for other third-generation sequencing technologies included both kinetic parameters and ordered nucleotide identities as an input data structure for a model (such as a recurrent neural network) to detect base modifications. EX1022, Abstract, 8-9, Fig. 1. As Dr. Dessimoz explains, POSAs appreciated that these workflows, while developed for ONT data, were applicable to identification of modified bases in SMRT data. *See* §I.A.3; EX1002, ¶86.

A POSA would have reasonably expected success in achieving the claimed invention as it requires no more than using pulse data generated in the course of SMRT sequencing in the manner taught by the Flusberg Publications.

Additionally, advances in computing power and deep learning networks that occurred in the years between Flusberg and the '794 patent further confirm a POSA would have reasonably expected success. Increased computing resources enabled routine and widespread use of data intensive models and processes, including in genomics. *E.g.*, EX1022, 8 (describing prediction of methylation based on 7 features provided by ONT sequencing); EX1038, 3-4 (discussing accuracy improvements from inclusion of multiple features). Moreover, while the claims are not limited to any particular model, the availability of open-source deep learning networks and the large number of models successfully applied to the

analysis of genomic data speak to the relative ease of implementation. *See* §I.A.4; *Intelligent Bio-Systems, Inc. v. Illumina Cambridge, Ltd.*, 821 F.3d 1359, 1367 (Fed. Cir. 2016) (“[O]ne must have a motivation to combine accompanied by a reasonable expectation of achieving what is claimed in the patent-at-issue.”). Specifically, the prior art at the time expressly embraced neural networks for detecting methylation status including the Keras framework used by PO. EX1017, 16 (cataloging deep learning resources, including Keras, for genomics applications); EX1035, 3, (“implemented by Keras framework”), 5 (describing “an intelligent and robust method” for methylation detection); EX1022, Fig. 1 (illustrating neural network for detecting methylation in ONT data); *see* EX1030, 10 (“The program for the CNN model was implemented on the basis of the Keras deep learning framework (<https://keras.io/>).”); EX1002, ¶¶87-90.

2. Claim 1

The Flusberg Publications and PacBio-BAM teach each element of claim 1, as discussed below.

i. Preamble

A method for detecting a modification of a nucleotide in a nucleic acid molecule, the method comprising;

To the extent the preamble is limiting, the prior art teaches such a method. For example, Flusberg-338 teaches “single molecule sequencing methods...for identification of modified nucleotides within nucleic acid sequences.” EX1007,

Abstract. Flusberg-338 teaches using “a classifier for distinguishing methylated from unmethylated cytosines in a nucleic acid template” based on a set of pulses’ features in a “real-time sequencing-by-synthesis trace.” EX1007, 66:46-67:4. Flusberg likewise teaches “direct detection of modified nucleotides in the DNA template” for various types of modification. EX1006, Abstract; *see id.*, 463-64; EX1011, 8 (“base modification detection”); EX1002, ¶¶91-93.

ii. Element 1.1

The prior art teaches Element 1.1. For example, Fig. 1 of Flusberg (reproduced below) shows a trace of fluorescence produced in an SMRT sequencing reaction reflecting each of the properties recited in Element 1.1, including nucleotide identity, position in the sample nucleic acid molecule, PW for each nucleotide, and IPD for each nucleotide.

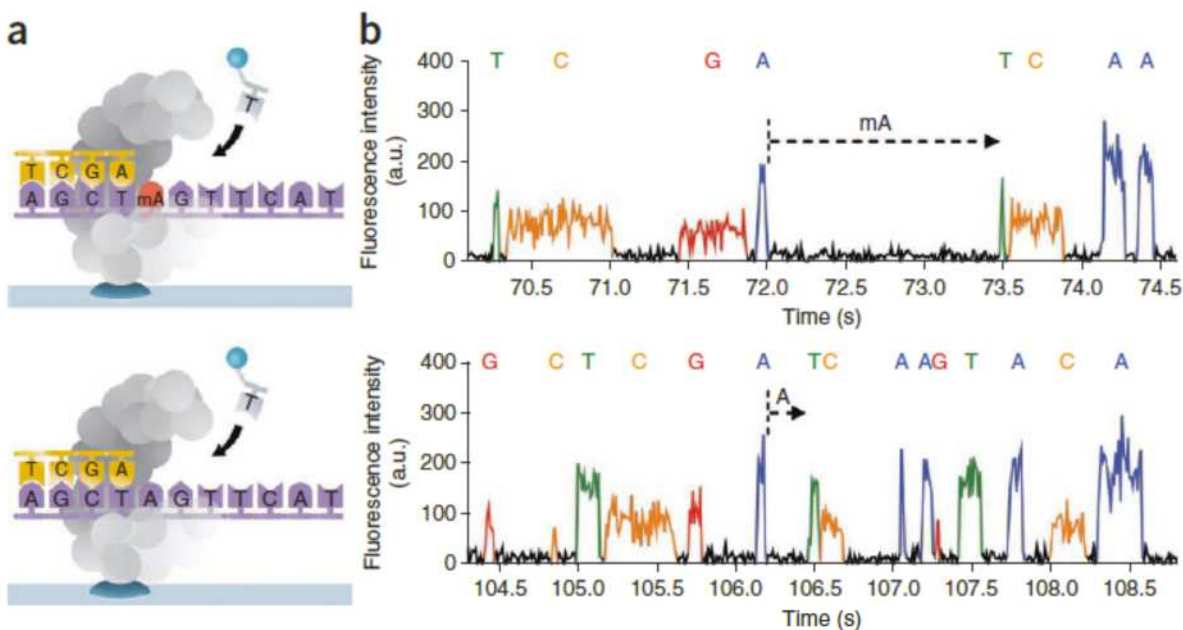


Figure 1 | Principle and corresponding example of detecting DNA methylation during SMRT sequencing. (a) Schematics of polymerase synthesis of DNA strands containing a methylated (top) or unmethylated (bottom) adenine. **(b)** Typical SMRT sequencing fluorescence traces for samples in **a**. Letters above the fluorescence trace pulses indicate the identity of the nucleotide incorporated into the growing complementary strand. Dashed arrows indicate the IPD before incorporation of the cognate thymine. For this typical example, the IPD is about five times larger for mA in the template compared to adenine.

EX1006, Fig. 1; EX1002, ¶¶94-95; *see* EX1014, 12-21 (relying on figure reproduced from Flusberg as disclosing obtaining values for each recited element).

Flusberg explains that each of the recited parameters are derived by measuring an optical signal corresponding to nucleotides in a sample nucleic acid molecule: a pulse’s “color identifies that nucleotide,” the order of nucleotides sequenced (shown by letters and colors in Fig. 1) determines the position of each nucleotide within the nucleic acid, and the pulses are further characterized by metrics including “pulse width and interpulse duration (IPD)” for each pulse. EX1006, 462. IPD is shown in Fig. 1 as a dashed line for a selected nucleotide in

each trace, and PW is reflected by the width of each colored pulse in Fig. 1.

EX1002, ¶95.

Flusberg-338 similarly describes SMRT sequencing which “distinctively identifies the base (e.g., A, G, T, or C)” as it is incorporated into a nascent strand, permitting “the practitioner to identify the base incorporated and, by complementarity, the sequence of the template DNA molecule.” EX1007, 58:32-44. The sequencing reaction data allows analysis of the kinetics based upon features of the pulses. *See* EX1007, 61:7-28, 61:46-54, 66:46-67:4. Flusberg-338 specifically identifies PW and IPD as being affected by the presence of nucleotide modifications, EX1007, 22:49-54, Figs. 2 (illustrating PW, IPD, and nucleotide identity and position measurable over a four-nucleotide sequence context), 16-19 (showing variation of PW and IPD from modifications). Flusberg-338 teaches that such pulse features may be measured for each nucleotide in a local context, and the set of measured values may be used to identify modifications using a model to classify nucleotides as modified or unmodified. *See* EX1007, 66:46-67:4; *see* EX1011 (BAM file containing each recited parameter); EX1002, ¶¶98-100.

iii. Element 1.2

(b) creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

an identity of the nucleotide,
a position of the nucleotide with respect to a target position
within the window,
the width of the pulse corresponding to the nucleotide, and
the interpulse duration;

The Flusberg Publications and PacBio-BAM teach Element 1.2. *See also* EX1037, 17-18, 36-39.

creating an input data structure comprising a window of nucleotides:

At the outset, the '794 patent treats the term “window” broadly, indicating a window may be symmetric or asymmetric around the target nucleotide, with unlimited variation in size. EX1001, 27:39-51, 33:12-48. Further, the measurement window need not be different from the set of all nucleotides in the data structure: “measurement windows may be the entire template molecule.” EX1001, 34:46-47. Thus, the recited “window” may be any set of nucleotides in a data structure—either all nucleotides of a sample nucleic acid molecule or any subset of nucleotides. EX1002, ¶¶101-02.

Flusberg-338 teaches “development of a classifier for distinguishing methylated from unmethylated cytosines in a nucleic acid template,” for which a “set of features is measured for every pulse” and used to form a “tabulation” of “local pulse features” over a “local context” around a “pulse of interest,” thereby forming a “feature vector.” EX1007, 66:46-67:4. Flusberg-338 suggests a

tabulation of the features of each pulse over a “local context extending 3 pulses to the left and right of the pulse of interest,” but notes that the size is variable. *Id.* As Dr. Dessimoz explains, the resulting feature vector is a “data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule,” corresponding to the claimed “input data structure” at least because the feature vector is an input for a model that distinguishes methylated and unmethylated nucleotides. EX1002, ¶¶103-04.

Regarding the recited window, Flusberg-338 describes tabulating a vector of k features for each of seven nucleotides—a center nucleotide (the “pulse of interest”) and three nucleotides on each side of the center nucleotide. EX1007, 66:46-67:4. This vector identifies the nucleotide pulse of interest by the label i and nucleotide pulses on each side by labels from $i-3$ to $i+3$, producing the following feature vector: $\vec{f} = \{f_1^{i-3}, f_2^{i-3}, \dots, f_k^{i-1}, f_1^i, f_2^i, \dots, f_k^i, \dots, f_1^{i+3}, f_2^{i+3}, \dots, f_k^{i+3}\}$. *Id.*; *see id.*, 64:29-65:20. As Dr. Dessimoz explains, Flusberg-338’s vector constitutes a window of nucleotides, with the example’s window size being seven. EX1002, ¶¶105-06.

Flusberg agrees, teaching “the kinetic signatures of methylation will be sequence context-dependent,” and proposing “detection of methylation may also be possible by tabulating expected kinetics over a suitable number of contexts or by taking advantage of heuristics that embody the observed trends in SMRT

sequencing kinetics.” EX1006, 465. By proposing tabulating kinetics over different contexts to detect methylation, Flusberg is teaching that the specific sequence of a window of nucleotides (the “context”) is an input to a model for detecting methylation, because the context affects kinetic signatures including PW and IPD. EX1002, ¶107.

In light of these teachings, a POSA would have created a data structure to store relevant features for a window of nucleotides. This structure could contain only features for the window, or it could store many nucleotides’ features with a specified subset being the window. EX1002, ¶108; *see* EX1010 (citing Flusberg) 130 (“seven bases upstream of and two bases downstream from the incorporation site”); *supra* §I.A.2.

Regarding the form of data structure used for the Flusberg Publications’ tabulated vectors, PacBio-BAM teaches creating an input data structure (a BAM file), which, as discussed below, contains each of the recited properties for a window of nucleotides. EX1002, ¶¶109-120; *see* EX1011, 1-2, 8 (BAM file properties used for methylation detection).

nucleotide identities: PacBio-BAM states that “PacBio-produced BAM files are fully compatible with the BAM specification.” EX1011, 1. The BAM specification referred to by PacBio-BAM teaches that in each BAM file, “[s]equence is encoded in 4-bit values, with adjacent bases packed into the same

byte starting with the highest 4 bits.” EX1012, 15. As Dr. Dessimoz explains, this encoding produces a data structure in the form of an array, with the bases A, C, G, and T identified as the 4-bit values 0001, 0010, 0100, and 1000, and the position of each base relative to each other base encoded by the order of the 4-bit values (i.e., the order of bases in the file corresponds to the order of bases in the nucleotide sequence). EX1002, ¶110.

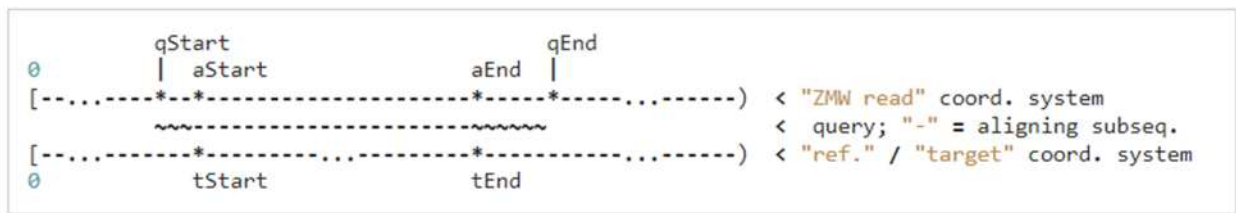
Including the identity of each nucleotide in the data structure, as taught by PacBio-BAM, is consistent with the Flusberg Publications’ teachings. Flusberg-338 teaches that its “feature vector” may include multiple “pulse features” for each nucleotide (EX1007, 66:46-67:4), and it also teaches that sequence context is a relevant “feature.” Flusberg-338 notes that “the local sequence context of the template strand can also influence and inhibit the activity of the polymerase along the template.” EX1007, 62:28-39; *see id.*, 64:53-55 (“It can also better integrate knowledge of the inter-pulse spacing and other information, such as sequence context, into a broad model.”), 66:4-45 (probability of methylation related to “neighboring sequence context” and features such as IPD), 5:33-46 (teaching a “combination of sequence context and methylation status affected interpulse duration” when detecting methylation). Indeed, Flusberg-338 expressly refers to basecalls as being a type of pulse feature used for a model making predictions indicating nucleotide modifications. *See id.*, 64:29-65:20 (prediction based on

“feature vector” containing “observed pulse data (both basecalls and other pulse features)” used in “model for identifying modified bases” relating to pauses), 7:4-17 (pause indicative of modification); EX1002, ¶111; §I.A.2.

nucleotide position with respect to a target position within the window:

PacBio-BAM includes the relative positions of each nucleotide, indicating where each nucleotide is in sequence, and thus relative to a nucleotide of interest in the window. As discussed above, PacBio-BAM stores nucleotide identities in an array, which provides the position of each nucleotide in sequence. EX1002, ¶112.

PacBio-BAM further teaches that the sequence may either be unaligned or aligned to a target sequence, such as a reference genome. EX1011, 1-2. The BAM file includes a position of the nucleotide with respect to a target position within a window. For example, PacBio-BAM describes aligning the start and end positions of a window of nucleotides (aStart, aEnd) within a sequence read (qStart, qEnd) to that of a target sequence (tStart, tEnd), as shown in the figure reproduced below:



EX1011, 1-2; *see* EX1014, 30 (relying on same figure); EX1002, ¶113.

Identifying the position of each nucleotide relative to a target nucleotide is also consistent with the Flusberg Publications. For example, Flusberg-338’s feature

vector labels each nucleotide's set of features with a superscript indicating its relative position, with superscript i representing the target nucleotide's pulse, and the other nucleotides being labeled from $i-3$ to $i+3$. EX1007, 66:46-67:4. These labels are direct indications of each nucleotide's position with respect to the target position (nucleotide i) within the window. EX1002, ¶114.

Flusberg similarly discloses tracking each nucleotide's relative position. Flusberg proposes "that methylated bases in a DNA template might be detected directly on the principle that their presence affects polymerase kinetics during SMRT sequencing." EX1006, 462. Flusberg teaches that PW and IPD "add valuable information about DNA polymerase kinetics" (*id.*) and that a combination of PW and IPD allows differentiation between unmodified cytosine, 5-methylcytosine, and 5-hydroxymethylcytosine. *See id.*, Fig. 3; EX1002, ¶115. To differentiate between these different types of unmodified and modified cytosines, Flusberg uses principal component analysis to find a group of weights for the respective PWs and IPDs of each of a set of nucleotides in a window. *See* EX1006, 462; *see* EX1001, claim 11 ("principal component analysis" as model). Like Flusberg-338, Flusberg describes that each weight is identified with a particular nucleotide based on its position relative to the nucleotide of interest, identified as nucleotide 0. *See* EX1006, 478; EX1002, ¶115.

PW and IPD: In addition to encoding the identity and relative locations of each nucleotide, PacBio-BAM encodes the PW and IPD of each nucleotide as integer arrays. EX1011, 8. PacBio-BAM confirms that these arrays are to be used for downstream analysis, noting that its encoding for IPD and PW is appropriate for “important production use cases,” and specifically for “base modification detection.” EX1011, 8.

Like the array of nucleotide identities, the PW and IPD arrays are data structures. EX1002, ¶¶116-17. As Dr. Dessimoz explains, a PacBio BAM file, by including arrays of sequenced nucleotide identities, PWs, and IPDs together, constitutes a data structure containing each property recited in Element 1.2 for each nucleotide in a window, which may be the set of all nucleotides (e.g., a BAM file storing only features for nucleotides in a window) or any subset of nucleotides (e.g., a BAM file containing many nucleotides’ features, with a subset specified as the window). EX1002, ¶117; *see* EX1014, 23-35 (asserting a BAM file generated on the PacBio SMRT platform includes values for each property of Element 1.2); EX1002, ¶¶69-73; §I.A.3; EX1038, 5-7 (extracting features for window of nucleotides for modification-detecting model input).

Including PW and IPD as features in an input data structure is further supported by the Flusberg Publications. Flusberg-338 teaches that the feature vector contains features measured for each pulse, including both PW and IPD. *See*

EX1007, 66:46-67:4. Flusberg-338 repeatedly confirms that PW and IPD are features that are affected by the presence of nucleotide modifications, both at the modification site and for nearby nucleotides. *See* EX1007, 9:41-46; 22:49-54; Figs. 12-13, 15-19 (showing PW and IPD ratios as a function of nucleotide position); EX1002, ¶118; §I.A.2.

Flusberg agrees. As discussed above, Flusberg teaches using PW and IPD to differentiate between unmodified cytosine, 5-methylcytosine, and 5-hydroxymethylcytosine by inputting them into a principal component analysis model for each nucleotide positioned around a target nucleotide in a window of nucleotides. *See* EX1006, 462, Fig. 3; EX1002, ¶119.

Accordingly, the prior art teaches creating an input data structure for detecting methylation that contains each of properties recited in Element 1.2 for each nucleotide within a window. EX1002, ¶120.

iv. Element 1.3

inputting the input data structure into a model, the model
trained by

The prior art teaches Element 1.3. For example, Flusberg-338 teaches “a classifier for distinguishing methylated from unmethylated cytosines in a nucleic acid template.” EX1007, 66:46-67:36. Flusberg-338 further discloses that “a generalized signal for determining methylation status on the trace” is derived by

measuring the “prior probabilities of methylated or un-methylated positions” and the “observed data likelihoods” of observing a feature vector given a methylated or unmethylated position. *Id.* As Dr. Dessimoz explains, Flusberg-338 is describing a model in which the feature vector $\vec{f}$, corresponding to the data structure discussed above in §IV.A.2.iii, would be an input. EX1002, ¶¶121-22.

Flusberg-338 discloses that this “general classifier...can be implemented with many standard statistical classification algorithms, i.e. linear discriminant analysis, multi-dimensional regression, kernel methods, classification and regression trees, neural networks, and support vector machines.” EX1007, 67:64-68:5. In the context of the ’794 patent claims, Flusberg’s classifier is a model. EX1002, ¶123. *Compare* EX1007, 67:67-68:5 (classifier implemented using “statistical classification algorithms” such as “neural networks, and support vector machines”), 65:13-16 (“Bayesian likelihood model for identifying modified bases”), *with* EX1001, 89:25-51 (models include various types of classifier, such as a “Bayes classifier,” a “neural network,” a “support vector machine,” etc.). A POSA would understand that models like Flusberg’s classifier would receive a data structure like that discussed in §IV.A.2.iii as an input and be trained via data structures containing the same elements. EX1002, ¶124; *see* EX1011, 8 (kinetic features in BAM file used for “base modification detection”); EX1021, 1-2 (trained model receiving BAM file as input).

Flusberg similarly teaches inputting an input data structure into a trained model. For example, it teaches using a principal component analysis model that takes a data structure including PW and IPD signals over a window of nucleotides as input and outputs a signal discriminating between different modification states (including unmodified) at a given nucleotide. EX1006, 463. This model includes weights for each nucleotide and pulse feature derived by fitting to training data. *Id.*, 478 (providing weights); EX1002, ¶¶125-27 (explaining training for principal component analysis). Flusberg further teaches that by using methods “to combine all available information,” its technique can be made applicable to variable sequence contexts. *See* EX1006, 464-65.

v. Element 1.4

Element 1.4 describes a training set for the model recited in Element 1.3. The Flusberg Publications and PacBio-BAM teach Element 1.4. According to Element 1.4, the training set must have a data structure for each of a plurality of nucleic acid molecules and a data structure that includes the properties recited in Element 1.2 for a window of nucleotides based on their measured pulses. Element 1.4 further requires the nucleotide in a target position to have a known modification state. The prior art teaches this element. EX1002, ¶128.

The Flusberg Publications teach measuring pulses to derive properties for a window of nucleotides in a plurality of nucleic acid molecules with known

modification states. *See* EX1006, Fig. 1 (showing pulse traces for methylated and unmethylated nucleic acids), 463-65 (describing analysis of properties of methylated and unmethylated nucleic acids), Fig. 3 (displaying data from pulse features from multiple nucleic acid molecules), 479 (providing different nucleic acid sequences), 478 (providing principal components extracted from pulse properties of methylated and unmethylated templates); EX1007, 66:46-67:36 (describing development of classifier on a training set comprising “methylated and unmethylated template data sets” analyzed over a window of features); *see id.*, 64:29-43, Fig. 20 (providing nucleotide sequences with identified modifications); EX1002, ¶129.

Further, a POSA would have known that to train a model to detect modifications from an input data structure, training data should be provided in the same format—and containing the same properties—as the input data structure. EX1002, ¶130. As discussed above for Element 1.2, PacBio-BAM and the Flusberg Publications teach an input data structure containing each property recited in Element 1.2, so to properly train a model for detecting nucleotide modifications, a POSA would receive a set of training data comprising data structures with the same contents derived from pulse characteristics from a window of nucleotides from a plurality of nucleic acid molecules with known modification states for target nucleotides in the molecules’ respective windows. EX1002, ¶130. The

Flusberg Publications and PacBio-BAM thus teach each limitation of Element 1.4.
EX1002, ¶131.

vi. Element 1.5

storing a plurality of first training samples, each including one
of the first plurality of first data structures and a first label
indicating the first state of the nucleotide at the target position

The prior art teaches Element 1.5. As discussed above for Elements 1.2 and 1.4, the prior art teaches creating and receiving the recited first data structures to be used as samples in a training set. A POSA would understand such training samples would be stored (e.g., as labeled PacBio BAM files) for training a model, as taught by the Flusberg Publications. EX1002, ¶¶132-34. For example, Flusberg-338 discloses providing “storage for the results of the methods of the invention.” EX1007, 73:35-42; *see, e.g., id.*, 66:46-67:36, 71:49-72:3, 73:6-25, 73:43-55; EX1002, ¶133.

A POSA would further understand that the training samples should be labeled to indicate the “first state of the nucleotide at the target position.” In other words, the training set should indicate whether a modification was present at the target position. Flusberg-338 teaches training a classifier based upon separate “methylated and unmethylated template data sets.” EX1007, 66:46-67:36. Both Flusberg Publications include labels indicating the presence of modified bases in sequences (*see* EX1007, Fig. 20; EX1006, 479), and a POSA would understand

that for a model to learn to detect modifications from a combination of modified and unmodified samples, the modification state of the nucleotide of interest must be specified. EX1002, ¶134. For example, Flusberg-338 teaches determining probabilities of obtaining particular feature vectors given methylation or non-methylation. *See* EX1007, 66:46-67:36. Determining these values requires identifying which samples are methylated. EX1002, ¶134. Likewise, producing the colored principal component graph shown in Fig. 3 of Flusberg requires identifying which data points have which modifications. *See* EX1006, Fig. 3; EX1002, ¶¶134-35. The Flusberg Publications and PacBio-BAM thus teach each limitation of Element 1.5.

vii. Element 1.6

optimizing, using the plurality of first training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels of the first labels when the first plurality of first data structures is input to the model, wherein an output of the model specifies whether the nucleotide at the target position in the respective window has the modification

Element 1.6 describes a well-known method of training a model, wherein the model's parameters are optimized so that the output (in this case, the target nucleotide being modified or not modified) best reproduces the labels of the training data. EX1002, ¶¶136-37. For the Flusberg Publications, the output specifies whether the nucleotide at the target position has a modification, such as a

methylation. *Id.* For example, Flusberg-338 discloses that by analyzing features of methylated and unmethylated samples, “a generalized signal for determining methylation status on the trace has been determined,” using a vector of features corresponding to the claimed data structure as an input. EX1007, 66:46-67:36; *see* EX1002, ¶137. Flusberg-338 further discloses “standard classification algorithm development techniques known to those of ordinary skill in the art may be applied to refine this approach,” teaching techniques including “linear discriminant analysis, multi-dimensional regression, kernel methods, classification and regression trees, neural networks, and support vector machines.” EX1007, 66:46-67:36, 67:64-68:5. As Dr. Dessimoz explains, these well-known approaches train models by optimizing the models’ parameters based on whether their output matches the labels of the training data. EX1002, ¶¶138-40.

viii. Element 1.7

(d) determining, using the model, whether the modification is present in a nucleotide at the target position within the window in the input data structure.

The Flusberg Publications and PacBio-BAM teach Element 1.7. As discussed above for Elements 1.2 and 1.3, the Flusberg Publications and PacBio-BAM teach inputting the recited input data structure having a window with a nucleotide at a target position. As discussed for Element 1.6, the Flusberg Publications further teach optimizing a model to detect methylation from an input

data structure. *See, e.g.*, EX1007, 66:46-67:36. Thus, Element 1.7 corresponds to using the trained model for its intended purpose: to determine whether a nucleotide at a target position is modified. EX1002, ¶¶141-44.

For the foregoing reasons, claim 1 would have been obvious over the Flusberg Publications and PacBio-BAM.

3. Claim 2

Claim 2 depends from claim 1 and further recites a series of limitations collectively encompassing performing Elements 1.2, 1.3, 1.4, and 1.7 on input data structures from a plurality of sample nucleic acid molecules obtained from a biological sample of a subject, each input data structure corresponding to a window of its respective sample nucleic acid molecule, wherein each of the data structures is received and input into the model, which determines whether a modification is present at a respective target location in the window of the input data structure. EX1002, ¶145.

The Flusberg Publications teach determining modifications on a plurality of nucleic acid molecules. *E.g.*, EX1007, 18:5-16 (discussing detecting “fluorescence signals from an array of thousands of ZMWs simultaneously”), 20:63-21:2 (discussing “highly parallel” analysis of entire methylome), 2:11-22; EX1006, Fig. 3 (showing methylation determination for multiple different nucleic acid molecules), 463 (discussing methylation analysis of multiple molecules). Indeed, a

POSA would understand that the purpose of developing a model like the ones taught by the Flusberg Publications is to detect modifications from multiple nucleic acid molecules. EX1002, ¶146. Doing so would provide the recited plurality of data structures, each having a window around a respective target nucleotide, and each producing a respective determination of modification at the target location. EX1002, ¶¶146-47.

Claim 2 would thus have been obvious over the Flusberg Publications and PacBio-BAM.

4. Claim 7

Claim 7 depends from claim 2 and further recites “wherein the modification is a methylation, the method further comprising: determining the modification is present at one or more nucleotides, and determining a clinically-relevant DNA fraction, a fetal methylation profile, a maternal methylation profile, a presence of an imprinting gene region, or a tissue of origin using the presence of the modification at one or more nucleotides.” The Flusberg Publications and PacBio-BAM teach these limitations.

The Flusberg Publications teach that the modification is a methylation and determining the modification is present at one or more nucleotides. *See, e.g.*, EX1006, Abstract (“direct detection of modified nucleotides in the DNA template, including n6-methyladenine, 5-methylcytosine and 5-hydroxymethylcytosine”);

EX1007, 66:46-67:4 (“a classifier for distinguishing methylated from unmethylated cytosines”), 3:47-53 (“detection of methylated bases within nucleic acid sequences using a real time direct detection of such methylated sites”); EX1002, ¶¶148-49; *supra* §IV.A.2.viii.

The Flusberg Publications further teach the remaining limitation of claim 7. For example, Flusberg-338 teaches that “methylation status differs between tissue types.” EX1007, 2:11-16. Thus, a POSA would understand detected methylation status could be used to distinguish between tissues of origin for a sample. EX1002, ¶150. For example, Flusberg-338 teaches that methylation has been “associated with cancer” and hydroxymethylation is an epigenetic marker “identified in certain cell types in the brain” (EX1007, 1:55-2:10). EX1002, ¶150. Further, Flusberg-338 teaches that methylation may distinguish which organism was the origin of tissue, e.g., distinguishing between organisms having “very similar” genomes “by association with the known methylation patterns” of the organisms., EX1007, 7:1-7; EX1002, ¶¶150-51.

Accordingly, claim 7 would have been obvious over the Flusberg Publications and PacBio-BAM.

5. Claim 10

Claim 10 depends from claim 1 and further requires “sequencing the sample nucleic acid molecule.” Sequencing nucleic acid molecules is discussed throughout

the Flusberg Publications. *E.g.*, EX1006, 461-63, 465; EX1007, 11:1-39, 60:45-67, 75:32-46; EX1002, ¶¶152-53. Accordingly, claim 10 would have been obvious over the Flusberg Publications and PacBio-BAM.

6. Claim 11

Claim 11 depends from claim 1 and further recites “the model includes a machine learning model, a principal component analysis, a convolutional neural network, or a logistic regression.” The ’794 patent specifies that “‘machine learning models’ may include models based on using sample data (e.g., training data) to make predictions on test data, and thus may include supervised learning,” and “may include statistical models.” EX1001, 16:34-40. The Flusberg Publications disclose this element. *See, e.g.*, EX1006, Fig. 3, 463, 466 (principal component analysis); EX1007, 63:22-27 (neural network), 64:28-58, 66:46-67:36 (machine learning, including “developing a classifier on a training set”), 67:64-68:5 (“standard statistical classification algorithms” including regression or neural network), 67:27-30 (principal component analysis), 69:36-47 (same); EX1002, ¶¶154-56 (convolutional neural network one of two primary types of neural network); EX1017, 1-2, 15. Accordingly, claim 11 would have been obvious over the Flusberg Publications and PacBio-BAM.

7. Claim 12

Claim 12 depends from claim 1 and further recites the window comprises nucleotides from first and second strands of the sample nucleic acid molecule, and the input data structure comprises a property value indicating which strand the nucleotide is on. The prior art teaches this element.

For example, Flusberg-338 teaches that a complementary pair of strands may be analyzed to detect methylation at a CG sequence that has a complementary CG sequence that may be jointly methylated. EX1007, 68:5-21. The complementary nucleotides on each strand would each be nucleotides in a window about the nucleotide of interest (i.e., the C at either of the complementary CG loci), and a POSA would understand that information identifying which nucleotides were on each strand would be needed to perform this analysis of complementary strands. EX1002, ¶¶157-59. Further, the BAM file according to PacBio-BAM identifies which nucleotides are on each complementary strand. *See* EX1011, 6 (indicating “forward” or “reverse” pass in “subread local context”).

Accordingly, claim 12 would have been obvious over the Flusberg Publications and PacBio-BAM.

8. Claims 14 and 17

Claim 14 depends from claim 1 and further recites “the nucleotides within the window are determined using a circular consensus sequence and without

alignment of the sequenced nucleotides to a reference genome.” Claim 17 recites identical language but specifies the determination is “without using a circular consensus sequence.” The Flusberg Publications and PacBio-BAM teach both of these claims.

Regarding alignment to a reference genome, PacBio-BAM teaches a BAM file may be either aligned or unaligned to a genome. EX1011, 1-2. Flusberg also discloses that synthetic sequences may be tested for modifications, and a POSA would understand that such sequences cannot be aligned to a reference genome. EX1006, 462, Fig. 4; *see* EX1007, 13:52-58, 75:33-46. A POSA would thus understand that while alignment to a genome may be helpful in some contexts, an unaligned sequence may be used in other contexts. EX1002, ¶¶160-61.

Regarding circular consensus sequencing, Flusberg-338 discloses that its modification detection can be performed on “double-stranded linear nucleic acid [or] double-stranded circular nucleic acid.” EX1007, 6:56-61; *see, e.g., id.*, 9:35-41 (redundant circular consensus sequencing optional), 28:53-57 (redundant sequencing not required for methylation detection). Flusberg-338 thus discloses that circular consensus sequencing is optional, but that non-circular or non-consensus sequencing may also be performed. EX1002, ¶162.

In conjunction, a POSA would have been motivated in appropriate circumstances to perform the process recited in claim 1 without alignment to a

reference genome, either with circular consensus sequencing as recited in claim 14 or without it as recited in claim 17. EX1002, ¶¶162-63.

Accordingly, claims 14 and 17 would have been obvious over the Flusberg Publications and PacBio-BAM.

9. Claim 15

Claim 15 depends from claim 1 and further recites “each nucleotide within the window is enriched or filtered.” Flusberg-338 teaches this element. *See* EX1007, 60:22-32, (“Enrichment strategies are beneficial.... [T]he resulting nucleic acid sample is less complex and more enriched for fragments of interest than was the original nucleic acid sample.”); *see id.*, 59:45-61:5; EX1002, ¶¶164-66.

Accordingly, claim 15 would have been obvious over the Flusberg Publications and PacBio-BAM.

10. Claim 18

Claim 18 depends from claim 1 and further recites “the optical signal is a fluorescence signal from a dye-labeled nucleotide.” The Flusberg Publications disclose this element. *See, e.g.*, EX1006, Abstract, 461-62, 466; EX1007, 21:42-45, 22:57-67; EX1002, ¶¶167-68.

Accordingly, claim 18 would have been obvious over the Flusberg Publications and PacBio-BAM.

11. Claim 19

Claim 19 depends from claim 1 and further recites “each window associated with the first plurality of data structures comprises 4 consecutive nucleotides on a first strand of each first nucleic acid molecule.” The Flusberg Publications teach this element.

For example, Flusberg-338 teaches that a window may vary in size, giving an example of a 7-nucleotide window. EX1007, 66:46-67:4 (“[A] local context extending 3 pulses to the left and right of the pulse of interest is assumed, but this context size is flexible....”). Flusberg likewise teaches using a window comprising 4 consecutive nucleotides. *See* EX1006, 478 (principal component weights computed 9-nucleotide window); EX1002, ¶¶169-72.

Accordingly, claim 19 would have been obvious over the Flusberg Publications and PacBio-BAM.

B. [Ground 2] Claims 3-6 and 9 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications, PacBio-BAM, and Clement

As discussed in detail below, the Flusberg Publications, PacBio-BAM, and Clement render claims 3-6 and 9 obvious. Claims 3-6 relate to classifications of a disorder such as cancer based on the presence of modifications determined according to claims 1 and 2. Claim 9 recites elements related to genome alignment and use of a cutoff number, both of which are steps useful when detecting the

disorders addressed in claims 3-6. The '794 patent defines "classification" as referring "to any number(s) or other characters(s) that are associated with a particular property of a sample." EX1001, 14:65-15:5. This includes binary classifications as well as more subtle classifications (such as classifications of different levels of disease). *See id.* As one example, cancer classifications for cancer may indicate a "level of cancer," which may include any of numerous characteristics, such as stage, size, metastasis, response to treatment, recurrence, or other measures. EX1001, 15:25-45; EX1002, ¶173.

The Flusberg Publications teach that "DNA methylation...has been associated with cancer and other disease processes," but it does not describe in detail how such disorders can be classified based on methylation patterns. EX1007, 1:56-61; *see* EX1006, 461. Clement teaches "methods of determining a cancer treatment prognosis" based on "evaluating epigenetic and genetic changes within a tumor sample." EX1018, Abstract. Such epigenetic evaluation includes classifying patterns of methylation. EX1018, ¶¶9, 78, 198. A POSA would have been motivated to classify cancer in a patient based on a pattern of methylation detected as taught by the Flusberg Publications and PacBio-BAM to aid in assessing cancer prognosis. EX1002, ¶174; *supra* §I.B.3.

A POSA would have a reasonable expectation of success in combining these references. The combination would be predictable and routine, as it would merely

use SMRT-detected methylation as taught by Flusberg and PacBio to classify cancer as taught by Clement. Each element would perform its ordinary function in the combination, with methylation patterns being detected as taught by the Flusberg Publications and PacBio-BAM, and those patterns being used as input data to classify cancer as taught by Clement. EX1002, ¶175. The combination would thus be combining elements according to known methods to yield predictable results. *Id.*

1. Claims 3-4

Claim 3 depends from claim 2 and further recites “determining the modification is present at one or more nucleotides” and “determining a classification of a disorder using the presence of the modification at one or more nucleotides.” Claim 4 depends from claim 3 and further recites “wherein the disorder comprises cancer.” These elements are taught by the Flusberg Publications, PacBio-BAM, and Clement.

As discussed above for claim 7, the Flusberg Publications teach determining whether a modification is present at one or more nucleotides, as recited in claim 3. *Supra*, §IV.B.1; EX1006, Abstract; EX1007, 3:47-53, 66:46-67:4; EX1002, ¶¶176-77.

The Flusberg Publications and Clement teach classifying a disorder comprising cancer using the presence of the modification. The Flusberg

Publications teach that changes in methylation patterns are associated with cancer and other diseases. EX1006, 461; EX1007, 1:56-61. Clement further teaches that by analyzing methylation patterns allows classification of a disease such as cancer. For example, Clement teaches that a methylation pattern for a subject may be classified as consistent (“concordant”) or inconsistent (“discordant”). EX1018, ¶78. Discordant or inconsistent methylation is associated with malignancy in a cancer patient, with inconsistent methylation over a sequence or across multiple genes indicating a “subject is more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof.” EX1018, ¶¶9, 198. Applying Clement in combination with the Flusberg Publications and PacBio-BAM, a POSA would classify a patient as likely having malignant cancer where a discordant methylation pattern was detected. EX1002, ¶¶178-79.

Accordingly, claims 3 and 4 would have been obvious over the Flusberg Publications, PacBio-BAM, and Clement.

2. Claim 5

Claim 5 depends from claim 4 and further recites “determining that the classification of the disorder is that the subject has the disorder, and treating the subject for the disorder by chemotherapy, radiation, or surgery.” These elements are taught by the Flusberg Publications, PacBio-BAM, and Clement.

As discussed above for claims 3 and 4, Clement teaches determining that a patient has “the disorder” (i.e., a malignant cancer more likely to develop resistance, relapse, or metastasize) based on a detected discordant methylation pattern. EX1018, ¶¶9, 198. Clement further teaches that in such a case, a patient may be treated with “surgery or radiation,” or by “altering treatment with an antitumor agent.” EX1018, ¶19; *see id.*, ¶70 (“The terms ‘antitumor agent’ and ‘chemotherapeutic agent’ are used interchangeably herein....”). Thus, Clement teaches treating the subject for the disorder with “chemotherapy, radiation, or surgery,” as recited in claim 5. EX1002, ¶¶180-82.

Accordingly, claim 5 would have been obvious over the Flusberg Publications, PacBio-BAM, and Clement.

3. Claim 6

Claim 6 depends from claim 3 and further recites “determining the classification of the disorder uses the number of modifications or the sites of the modifications.” Clement teaches this element.

Clement teaches that discordant methylation patterns indicate likely malignant cancer, whereas concordant patterns indicate less likely malignancy. EX1018, ¶¶9, 198. A discordant pattern is one where methylation is “locally disordered,” in contrast to a “concordant” pattern in which “all CpGs contained within one sequencing read are uniformly methylated or uniformly unmethylated.”

EX1018, ¶78. The concordant/discordant determination relates to the number or sites of modifications because it compares the number of methylations in a local sequence region to the number of CpG sites, with concordance corresponding to all or no CpG sites in the region methylated and discordant corresponding to some but not all local CpG sites methylated. EX1002, ¶¶183-85.

Accordingly, claim 6 would have been obvious over the Flusberg Publications, PacBio-BAM, and Clement.

4. Claim 9

Claim 9 depends from claim 2 and further recites “the plurality of sample nucleic acid molecules align to a plurality of genomic regions,” for each of which “a number of sample nucleic acid molecules is aligned to the genomic region,” and “the number of sample nucleic acid molecules is greater than a cutoff number.” The Flusberg Publications, PacBio-BAM, and Clement teach these elements.

Each of the Flusberg Publications and PacBio-BAM teaches that the sequences in a BAM file may be aligned to a reference genome. EX1011, 1-2; EX1006, 461 (“align to the reference genome”), 466 (alignment using “Smith-Waterman algorithm”); EX1007, 50:30-33 (discussing “align[ing] and map[ping] to a known sequence, e.g., from a genomic sequence database”), 23:32-36 (“comparison to a reference sequence”); *see* EX1007, 2:11-22, 13:24-40; EX1006, 461; EX1002, ¶¶186-87.

Clement similarly teaches aligning to a reference genome, either “across the entire genome or for a specified genomic region,” and it further teaches evaluating CpG values “within the region of interest with read depth greater than 10.” EX1018, ¶235. This involves determining whether a pattern of methylation detected in a set of reads constitutes a concordant or discordant pattern, as illustrated below in Fig. 3E:

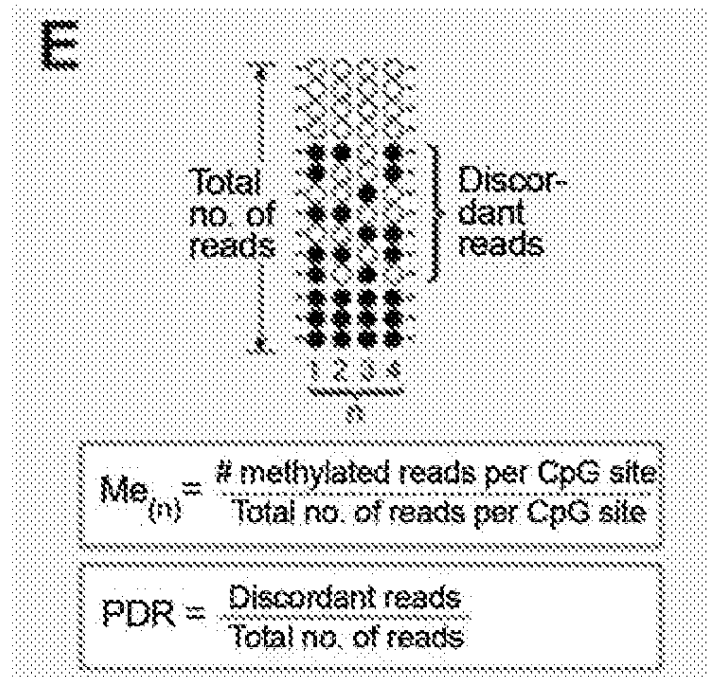


FIG. 3E

EX1018, Fig. 3E. Clement’s employment of a read depth limit constitutes a cutoff number requiring the number of nucleic acids overlapping at a CpG site at a

genomic region to exceed the cutoff. EX1002, ¶¶188-89; *see* EX1018, ¶¶55-56, 58, 249 (applying various alternative cutoff values).

A POSA would have been motivated to use a cutoff number to the number of nucleic acid molecules aligning to a genomic region, as taught by Clement, as doing so would help distinguish real patterns from noise. EX1002, ¶¶190-91. For example, in Fig. 3E above, determining that the labeled set of reads is discordant can be made more confident by increasing the number of reads needed beyond a cutoff number. *Id.*

Accordingly, claim 9 would have been obvious over the Flusberg Publications, PacBio-BAM, and Clement.

C. [Ground 3] Claims 8 and 15-16 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications, PacBio-BAM, and the PacBio Template Guide

As discussed in detail below, the Flusberg Publications, PacBio-BAM, and the PacBio Template Guide render claims 8 and 15-16 obvious. Claims 8 and 15-16 include limitations related to filtering DNA molecules, such as by size selection. The Flusberg Publications teach using PacBio's SMRT sequencing to detect methylation. EX1006, Abstract; EX1007, 18:5-16. The PacBio Template Guide teaches the standard procedure for preparing DNA templates for SMRT sequencing. EX1019, 2; EX1002, ¶192. These procedures include size selection steps. *E.g.*, EX1019, 24, 35. A POSA would have had reason to filter nucleic acid

molecules using the size selection processes in the PacBio Template Guide because it teaches routine procedures to prepare molecules for PacBio sequencing and it “highly recommends” such size selection. EX1019, 35; EX1002, ¶192; *supra* §I.B.4.

A POSA would have a reasonable expectation of success in combining these references. A POSA would have understood that PacBio’s SMRT sequencing would typically be performed in conjunction with the template preparation steps, such as size-selection, that the PacBio Template Guide teaches. EX1002, ¶193. Thus, the combination would merely involve combining elements according to known methods to yield routine, predictable results. *Id.*

1. Claim 8

Claim 8 depends from claim 2 and further recites “each sample nucleic acid molecule of the plurality of sample nucleic acid molecules has a size greater than a cutoff size.” The ’794 patent discloses that a “cutoff” refers to a “predetermined number,” and, for example, “a cutoff size can refer to a size above which fragments are excluded.” EX1001, 15:6-8. Claim 8 involves a cutoff in the opposite direction, with a plurality of nucleic acid molecules being above a given size. EX1002, ¶194; *cf.* EX1001, 91:32-42. This element is taught by the PacBio Template Guide.

The PacBio Template Guide teaches a process to prepare nucleic acid templates such as double-stranded DNA molecules to be used in PacBio's SMRT sequencing. *See* EX1019, 3. The Template Guide teaches that it is "highly recommended to perform a size-selection step" when sequencing large libraries to avoid loading bias in ZMW wells. EX1019, 35. In one example, size selection is performed with AMPure beads to remove "ligation products smaller than 0.4 kb." EX1019, 24. This size selection step excludes nucleic acid molecules below a cutoff size (e.g., 400 bases), leaving a remaining plurality of nucleic acid molecules each having a size above the cutoff. EX1002, ¶¶195-96. This step thus meets the limitations of claim 8. *Id.*; *cf.* EX1004, 200 (arguing such exclusion meets claim 8's limitations); EX1014, 73-78 (asserting PacBio size selection meets these limitations).

Accordingly, claim 8 would have been obvious over the Flusberg Publications, PacBio-BAM, and the PacBio Template Guide.

2. Claims 15-16

Claim 15 depends from claim 1 and further recites "wherein each nucleotide within the window is enriched or filtered." Claim 16 depends from claim 15 and further characterizes how the nucleotides are either enriched or filtered. Regarding filtering, claim 16 specifies that the nucleotides are "filtered by: selecting double-

stranded DNA molecules having a size with a size range.” The PacBio Template Guide teaches these limitations.

As discussed above for claim 8, the PacBio Template Guide teaches size selection for double-stranded DNA molecules. *See* EX1019, 3, 24, 35. Selecting double-stranded DNA molecules with sizes above a cutoff constitutes selecting double-stranded DNA molecules having a size with a size range; specifically, the size range is greater than the cutoff. EX1002, ¶¶197-99. The Template Guide further teaches that it is “highly recommended to select amplicons of the same size (+/-10%) to minimize loading bias.” EX1019, 35. Selecting amplicons (i.e., double stranded DNA molecules) within a +/-10% size range would similarly meet the filtering and size range limitations of claims 15 and 16.

Accordingly, claims 15 and 16 would have been obvious over the Flusberg Publications, PacBio-BAM, and the PacBio Template Guide.

D. [Ground 4] Claims 13 and 15-16 are unpatentable as obvious under 35 U.S.C. §103 over the Flusberg Publications, PacBio-BAM, and Tsai

As discussed in detail below, the Flusberg Publications, PacBio-BAM, and Tsai render claims 13 and 15-16 obvious. Claims 13 and 15-16 recite limitations related to enriching nucleotides by using a Cas9 complex and hairpin adapters to form circular DNA molecules. The Flusberg Publications teach ligating hairpin adapters to form circular DNA, but they do not expressly teach cutting DNA using

a Cas9 complex. EX1007, 60:54-59 (teaching use of hairpin adapters for enrichment), EX1006, 463 (teaching ligation of hairpin adapters). Tsai is a PacBio patent publication teaching cutting double-stranded DNA templates with a Cas9 complex and ligating hairpin adapters to form circular DNA molecules for SMRT sequencing. EX1020, ¶¶7, 41. Tsai further teaches using its techniques to prepare DNA for base modification detection “by monitoring the kinetics of the sequencing reaction,” citing to the PCT publication of Flusberg (WO2012/065043, EX1024). EX1020, ¶23; *supra* §I.B.5. Tsai teaches that using Cas9 complexes to prepare DNA templates for sequencing allows targeting “regions having sites that may have epigenetic modifications” related to various disorders. EX1020, ¶102. A POSA would thus have been motivated to combine Tsai with the Flusberg Publications and PacBio-BAM to obtain the benefits Tsai suggests from such a combination. EX1002, ¶200. A POSA would also have a reasonable expectation of success in such a combination, as each reference relates to the same DNA sequencing technology, combined according to known methods, and yielding routine, predictable results. *Id.*

1. Claim 13

Claim 13 depends from claim 12 and further recites “the sample nucleic acid molecule is a circular DNA molecule formed by cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and

ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule.”

These elements are taught by Tsai.

Tsai teaches exposing “guide RNAs to Cas9 endonucleases...to form a sgRNA-Cas9 complex,” binding the Cas9 complex to a nucleic acid sample and causing “the Cas9 endonucleases [to] cleave the nucleic acid sample.” EX1020, ¶7.

Tsai further teaches “linking stem-loop adapters to all double-stranded ends in the nucleic acid sample after the Cas9 cleavage to form adapter-ligated fragments.”

Id.; *see id.*, ¶55. Tsai’s process is shown in Fig. 8:

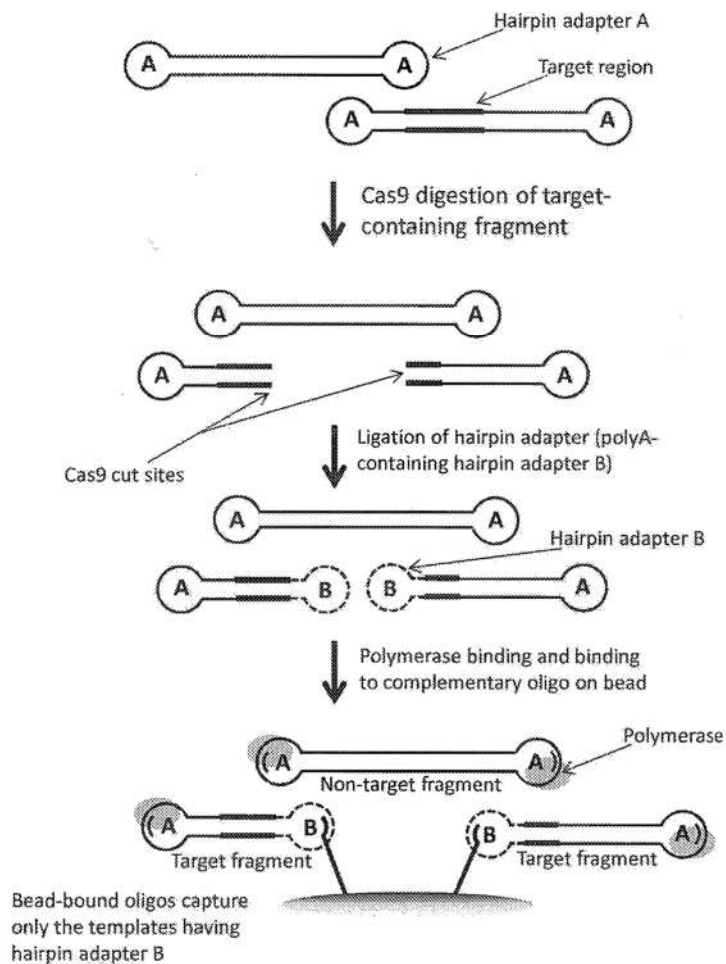


Figure 8

EX1020, Fig. 8; EX1002, ¶¶201-02.

Accordingly, claim 13 would have been obvious over the Flusberg Publications, PacBio-BAM, and Tsai. EX1002, ¶¶202-03.

2. Claims 15-16

Claim 15 depends from claim 1 and further recites “wherein each nucleotide within the window is enriched or filtered.” Claim 16 depends from claim 15 and further characterizes how the nucleotides are either enriched or filtered. Regarding

enrichment, claim 16 specifies that the nucleotides are enriched by the same steps recited for claim 13. As discussed above, Tsai teaches performing the steps recited in claim 13. Accordingly, for the same reasons, Tsai teaches enriching as recited in claims 15 and 16.

For the foregoing reasons, claims 13, 15, and 16 would have been obvious over the Flusberg Publications, PacBio-BAM, and Tsai. EX1002, ¶¶204-07.

V. THERE ARE NO *FINTIV* ISSUES

This petition does not implicate the Board's discretion according to *Fintiv*. *Apple Inc., v. Fintiv, Inc.*, IPR2020-00019, Paper 11. PO filed its district court complaint in Delaware on December 14, 2022. The court granted a motion to transfer the case to the Northern District of California on August 2, 2023 and the case was transferred on August 16, 2023. As of the filing of this petition, a trial date has not been set but is reasonably expected to be after issuance of a final written decision here.

VI. GROUNDS FOR STANDING

Petitioner certifies that, under 37 C.F.R. §42.104(a), the '794 patent is available for *inter partes* review, and Petitioner is not barred or estopped from requesting *inter partes* review of the '794 patent on the grounds identified.

VII. CONCLUSION

For the reasons set forth above, claims 1-19 of the '794 patent are unpatentable. Petitioner therefore requests that a *inter partes* review of these claims be instituted.

Respectfully submitted,

Dated: October 17, 2023

/Michael T. Rosato/
Michael T. Rosato, Lead Counsel
Reg. No. 52,182

VIII. MANDATORY NOTICES UNDER 37 C.F.R. §42.8

Real Party-in-Interest (37 C.F.R. §42.8(b)(1)): Pacific Biosciences of California, Inc. is the real party-in-interest.

Related Matters (37 C.F.R. §42.8(b)(2)): Patent Owner has asserted the '794 patent against Petitioner in the United States District Court in Northern California (5:23-cv-04166).

Lead and Back-Up Counsel (37 C.F.R. §42.8(b)(3))

Lead Counsel: Michael T. Rosato (Reg. No. 52,182)

Back-Up Counsel: Sonja R. Gerrard (Reg. No. 72,802); Patrick M. Medley (Reg. No. 74,735)

Service Information – 37 C.F.R. §42.8(b)(4). Petitioner hereby consents to electronic service.

Email: mrosato@wsgr.com; sgerrard@wsgr.com; pmedley@wsgr.com

Post: Wilson Sonsini Goodrich & Rosati, 701 Fifth Avenue, Suite 5100,
Seattle, WA 98104-7036

Tel.: 206-883-2529

Fax: 206-883-2699

IX. CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. §42.24(d), the undersigned certifies that this Petition complies with the type-volume limitation of 37 C.F.R. §42.24(a). The word count application of the word processing program used to prepare this Petition indicates

that the Petition contains 13,985 words, excluding the parts of the brief exempted by 37 C.F.R. §42.24(a).

Dated: October 17, 2023

/Michael T. Rosato/
Michael T. Rosato, Lead Counsel
Reg. No. 52,182

X. PAYMENT OF FEES UNDER 37 C.F.R. §§42.15(A) AND 42.103

The required fees are submitted herewith. If any additional fees are due at any time during this proceeding, the Office is authorized to charge such fees to Deposit Account No. 23-2415.

XI. APPENDIX – LIST OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 11,091,794 to Lo (“the ’794 patent”)
1002	Declaration of Dr. Christophe Dessimoz
1003	Curriculum Vitae of Christophe Dessimoz, Ph.D.
1004	U.S. Patent Application No. 16/995,607 File History
1005	U.S. Patent Application No. 17/379,544 File History
1006	Flusberg, et al., “Direct Detection of DNA Methylation During Single-molecule, Real-Time Sequencing,” <i>Nature Methods</i> 7(6) (2010) (“Flusberg”)
1007	U.S. Patent No. 9,175,338 to Flusberg et al. (“Flusberg-338”)
1008	Clark, et al. “Direct Detection and Sequencing of Damaged DNA Bases” <i>Genome Integrity</i> 2 (2011) (“Clark”)
1009	Loomis et al., “Sequencing the Unsequenceable: Expanded CGG-repeat Alleles of the Fragile X Gene,” <i>Genome Research</i> 23(1) (2013) (“Loomis”)
1010	Schadt, et al., “Modeling Kinetic Rate Variation in Third Generation DNA Sequencing Data to Detect Putative Modifications to DNA Bases,” <i>Genome Research</i> 23(1) (2013) (“Schadt”)
1011	BAM Format Specification for PacBio, PacBioFileFormats 5.0 Documentation, http://web.archive.org/web/20170916112547/http://pacbiofileformats.readthedocs.io/en/5.0/BAM.html (archived Sep. 16, 2017) (“PacBio-BAM”)
1012	SAM/BAM Specification Working Group, Sequence Alignment/Map Format Specification, http://web.archive.org/web/20180613134937/http://samtools.github.io/hts-specs/SAMv1.pdf (archived Jun. 13, 2018) (“SAM/BAM Specification”)

1013	Davis et al., “Entering the Era of Bacterial Epigenomics with Single Molecule Real Time DNA Sequencing,” <i>Current Opinion in Microbiology</i> , 16(2) (2013) (“Davis”)
1014	<i>Take2 Tech, Ltd. v. Pac. Biosciences of Cal., Inc.</i> , DDE-1-22-cv-01595, Infringement Claim Chart for U.S. Patent No. 11,091,794 (“Take2 Infringement Contentions”)
1015	Feng, et al, “Detecting DNA modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic,” <i>PLoS Computational Biology</i> , 9(3) (2013) (“Feng”)
1016	“Detecting DNA Base Modifications Using Single Molecule, Real-Time Sequencing,” https://www.pacb.com/wp-content/uploads/2015/09/WP_Detecting_DNA_Base_Modifications_Using_SMRT_Sequencing.pdf (2012) (“PacBio White Paper”)
1017	Zou et al., “A Primer on Deep Learning in Genomics” <i>Nature Genetics</i> , 51(1) (Jan. 2019)
1018	U.S. Patent Pub. No. 2016/0326593 to Clement et al. (“Clement”)
1019	Pacific Biosciences Template Preparation and Sequencing Guide, https://www.pacb.com/wp-content/uploads/2015/09/Guide-Pacific-Biosciences-Template-Preparation-and-Sequencing.pdf (2015) (“PacBio Template Guide”)
1020	U.S. Patent Publication No. 2016/0208241 to Tsai et al. (“Tsai”)
1021	Pacific Biosciences kineticsTools Manual, http://web.archive.org/web/20170813085437/https://github.com/PacificBiosciences/kineticsTools/blob/master/doc/manual.rst (2018) (“kineticsTools manual”)
1022	Liu et al., “Detection of DNA Base Modifications by Deep Recurrent Neural Network on Oxford Nanopore Sequencing Data,” <i>Nature Comms.</i> 10(1) (Jun. 4, 2019) (“Q. Liu”)
1023	Eid et al., “Real-time DNA sequencing from single polymerase molecules,” <i>Science</i> 323(5910) (2009)

1024	Int. Appl. Publ. No. WO 2012/065043 (“Flusberg PCT”)
1025	Jonas Korlach & Stephen W. Turner, “Single-Molecule Sequencing” in <i>Encyclopedia of Biophysics</i> (2013) (“Korlach”)
1026	GenomeWeb, “PacBio Reveals Beta System Specs for RS; Says Commercial Release is on Track for First Half of 2011,” https://www.genomeweb.com/sequencing/pacbios-q3-losses-widen-it-readies-launch-rs-system (2010) (“GenomeWeb”)
1027	Jacob Schreiber et al., “Error Rates for Nanopore Discrimination among Cytosine, Methylcytosine, and Hydroxymethylcytosine along Individual DNA Strands,” <i>Proceedings of the National Academy of Sciences</i> 110(47) (2013) (“Schreiber”)
1028	Langlotz, A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop, <i>Radiology</i> 291(3) (Apr. 16, 2019) (“Langlotz”)
1029	Miotto et al., “Deep Learning for Healthcare: Review, Opportunities and Challenges,” <i>Briefings in Bioinformatics</i> , 19(6) (2018) (“Miotto”)
1030	Tse et al., “Genome-wide Detection of Cytosine Methylation by Single Molecule Real-time Sequencing,” <i>Proceedings of the National Academy of Sciences</i> , 118(5) (2021) (“Tse”)
1031	Tiernan Ray, “AI is Changing the Entire Nature of Compute,” <i>ZDNet</i> , http://web.archive.org/web/20190630155704/https://www.zdnet.com/article/ai-is-changing-the-entire-nature-of-compute/ (June 30, 2019) (“Ray”)
1032	Wenger et al., “Accurate Circular Consensus Long-read Sequencing Improves Variant Detection and Assembly of a Human Genome,” <i>Nature Biotechnology</i> , 37(10) (Aug. 12, 2019) (“Wenger”)
1033	Pacific Biosciences, PacBioFileFormats Documentation (Feb. 2022) (“PacBioFileFormats-2022”)
1034	Lau et al., “LongISLND: <i>in silico</i> Sequencing of Lengthy and Noisy Datatypes,” <i>Bioinformatics</i> , 32(24) (2016) (“Lau”)

1035	Tahir et al., “iRNA-PseKNC(2methyl): Identify RNA 2’-O-methylation Sites by Convolution Neural Network and Chou’s Pseudo Components,” <i>Journal of Theoretical Biology</i> , 465 (2018) (“Tahir”)
1036	DeepVariant Usage Guide, https://github.com/google/deepvariant/blob/42b9c777436d8306f6cb9d2e18da756fa370fcd8/docs/deepvariant-details.md (“DeepVariant”)
1037	U.S. Patent Application No. 18/189,084 File History
1038	Liu et al., “Accurate Detection of m6A RNA Modifications in Native RNA Sequences,” https://www.biorxiv.org/content/10.1101/525741v1 (Jan. 21, 2019) (“H. Liu”)
1039	First Declaration of Nathaniel E. Frank-White
1040	Second Declaration of Nathaniel E. Frank-White
1041	Declaration of Mozhgan Novbakhtian
1042	GitHub, PacBio BAM 3.0, https://github.com/PacificBiosciences/PacBioFileFormats/blob/3.0/BAM.rst (Mar. 28, 2016)
1043	GitHub, PacBio BAM 5.0, https://github.com/PacificBiosciences/PacBioFileFormats/blob/5.0/BAM.rst (Mar. 7, 2017)
1044	PacBio Sequencing – How it Works

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§42.6(e) and 42.105(a), this is to certify that I caused to be served a true and correct copy of the foregoing Petition for Inter Partes Review of U.S. Patent No. 11,091,794 (and accompanying Exhibits 1001-1044) by overnight courier (Federal Express or UPS), on this 17th day of October 2023, on the Patent Owner at the correspondence address of the Patent Owner as follows:

Kilpatrick Townsend & Stockton LLP
MailStop: IP Docketing-22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

Respectfully submitted,

Dated: October 17, 2023

/Michael T. Rosato/
Michael T. Rosato, Lead Counsel
Reg. No. 52,182