

Paper No. _____
Filed: October 28, 2024

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

PETITIONER'S REPLY
U.S. PATENT NO. 11,091,794

TABLE OF CONTENTS

	<u>Page</u>
I. Introduction.....	1
II. The Prior Art Teaches or Suggests the Claimed “input data structure”	2
A. PO fails to distinguish its “input data structure” from PacBio’s BAM file.....	2
i. PO advances the same argument rejected at institution.....	2
ii. PO’s arguments are not commensurate with the claims.....	4
B. The Prior Art Does not Teach Away From “identity of the nucleotide”	7
i. The Flusberg Publications teach nucleotide identity	7
ii. Flusberg’s experimental example uses nucleotide identity	10
iii. Prior art corroborates obviousness.....	11
C. The Prior Art Does not Teach Away From PW	14
III. No Objective Indicia Supporting Nonobviousness	16
A. Bisulfite sequencing and modification.....	18
B. Using a prior art model according to prior art teachings is obvious	21
C. No unexpected results	25
IV. Dependent Claims.....	27
V. Conclusion	27
VI. Certificate of Compliance.....	28
VII. Appendix – Petitioner’s Updated Exhibit List	29

I. INTRODUCTION

Aside from the “input data structure”—Element 1.2—Patent Owner (“PO”) does not challenge the sufficiency of the petition. Nor does PO dispute motivation to combine the Flusberg Publications and PacBio-BAM. Instead, PO’s challenge to prima facie obviousness focuses exclusively on arguing the claimed “input data structure” is not obvious. But PO’s arguments amount to nothing more than a piecemeal attack on the asserted references.

For PacBio-BAM, PO ignores the obviousness theory of the petition and the broad scope of its claims to argue that the BAM file cannot satisfy the input data structure. The Board rejected this argument at institution and, to the extent the record has changed, it is not in PO’s favor. PO acknowledges that PacBio’s BAM file contains SMRT sequencing reaction data for all properties recited in the claim. POPR, 13. PacBio-BAM specifically suggests using the saved information to detect base modifications. EX1011, 8. As to the Flusberg Publications, PO ignores their salient teachings that all recited properties were useful for detecting base modifications. PO then compounds its blinkered approach by pretending that the disclosed examples somehow teach away from the claimed input data structure. PO’s teaching away argument depends on mischaracterizing and ignoring the teachings of the references and is ultimately contrary to the evidence.

PO's secondary considerations arguments fare no better. PO argues objective indicia relating to direct detection of 5mC without bisulfite conversion. But PO's claims do not require 5mC detection nor do they exclude bisulfite conversion. Moreover, direct detection of base modifications (including 5mC) without bisulfite conversion is described in the Flusberg Publications, amongst several other PacBio publications.

Claims 1-19 of the '794 patent should be found unpatentable.

II. THE PRIOR ART TEACHES OR SUGGESTS THE CLAIMED "INPUT DATA STRUCTURE"

A. PO fails to distinguish its "input data structure" from PacBio's BAM file.

PO rehashes its preliminary response argument that PacBio's BAM file cannot satisfy the input data structure recited in the claim but fails to address any of the reasons the Board rejected PO's arguments at institution. Worse yet for PO, its expert has confirmed the broad scope of its claims, making plain that the claimed "input data structure" encompasses PacBio's BAM file.

i. PO advances the same argument rejected at institution

The petition established that: (1) the claimed properties—sequence context (identity and position) and polymerase kinetic parameters (PW & IPD)—were conventionally generated during SMRT sequencing and stored in PacBio's BAM file in a manner compatible with base modification detection; and (2) a POSA

would have found it obvious to consider the claimed properties together when detecting base modifications.

PO argues (p. 20) that PacBio's BAM file cannot satisfy the input data structure of Element 1.2. According to PO and Dr. Koo, the BAM file is a compressed binary format and "it comprises text data, not an input data array of numerical values" and "includes extraneous information...that needs to be parsed out before the data could be used by a trainable model." EX2041, ¶¶129-130. This argument is not persuasive and fails to undermine any of the Board's sound reasons for rejecting this same argument at institution. DI, 21-22.

PO's argument falsely assumes that a POSA would thoughtlessly include text or other extraneous data to impair the utility of the input data structure. But a POSA is neither a saboteur nor an "automaton." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 421 (2007).

Conversely, PacBio-BAM expressly discloses that PacBio's BAM file is compatible with downstream analyses, including base modification detection. EX1011, 8. PO provides no explanation for why a POSA would be unable to decompress or parse a BAM file (to the extent such operations are necessary) for claim 1. Plainly, a POSA would have done so. As Dr. Dessimoz explains, such operations would have been "trivial" and "any person of ordinary skill in the art would be able to do." EX2037, 103:2-6; *see also id.*, 102:11-20

(“...straightforward operation...”), 103:17-21 (“...trivial...”), 104:2-13 (“...This is really a very trivial step.”).

Moreover, Dr. Dessimoz’s unchallenged testimony is corroborated by the prior art. For example, the kineticsTools model takes as input a BAM file and uses the information stored in that file to detect base modifications. EX1021, 2-3. Thus, even if the recited properties are to be parsed from a decompressed BAM file, doing so was routine and well within the skills of a POSA. PO provides no argument or evidence to the contrary.

ii. PO’s arguments are not commensurate with the claims

PO’s “input data structure” arguments also fail because they are not commensurate with the broad scope of the claims. For example, Dr. Koo testified that, in computer science, data structures are simply “objects that contain [] information.” EX1049, 132:21-25. PO argues the BAM file cannot be directly read by a trainable model, but the claims recite no such requirement. PO’s claims generically recite a “model,” and the specification confirms the staggering breadth of the term “model.” PO does not propose a limiting claim construction for “model.” Nor could it. The patent places essentially no restriction on models that may be used within the scope of the invention and expressly permits inclusion of sub-modules.

The intrinsic evidence repeatedly confirms that the claim term “model” is both broad and generic. For example, claim 11 recites “wherein the model [of claim 1] **includes** a machine learning model, a principal component analysis, a convolutional neural network, or a logistic regression.” By specifying that the “model” can “include” a model, claim 11 confirms that the “model” in claim 1 need not be singular.

The specification similarly emphasizes the breadth of the term “model” as used in the claim. EX1001, 26:34-45 (“any model or data analysis framework can be developed”), 29:39-43, 72:53-55, 77:39-67, 89:5-67; *see also id.*, 16:41-46 (“The term ‘data analysis framework’ may include algorithms and/or models that can take data as an input and then output a predicted result.”); EX1049, 40:3-42:11. When discussing machine learning models, the specification explains that they are often developed using a computer and describes them broadly and generically. EX1001, 16:34-40. Accordingly, the claim language and specification do not preclude the model from employing a module to decompress or parse the reaction data.

PO argues (p. 20) that the BAM file cannot be an input data structure because “extraneous” information is present in the file. EX2041, ¶¶129-130. But Element 1.2 is open-ended; it recites “the input data structure comprising....” EX1049, 73:12-75:8. There is no dispute that PacBio’s BAM file contains the

required properties for input purposes. It is irrelevant if the BAM file contains additional information.

PO repeatedly argues (pp. 2-3, 13-20) the prior art held nucleotide identity constant rather than making it a variable. PO's argument is wrong and fails because Element 1.2 does not require "identity of the nucleotide" to be a variable and is agnostic as to whether the sample molecule's sequence was known/unknown in advance; it merely requires values for this property be embodied in a data structure for input. Furthermore, PO does not dispute that the prior art describes receiving values for "identity of the nucleotide" as required by Element 1.1—nor could it. *Compare* EX1014, 12-14, *with* EX1006, Fig. 1. PO also does not dispute that—variable or not—values for "identity of the nucleotide" are stored in PacBio's BAM file. EX1014, 23-36.

Ignoring the obviousness combination, PO attacks (pp. 18-19) PacBio-BAM individually as not disclosing a method for base modification detection. But PO challenges only Element 1.2. Even if PO's argument is accepted (it should not be), it does not address the obviousness of creating from the BAM file an input data structure comprising sequence context, IPD, and PW in view of the prior art as argued in the petition (Pet., 37-44). PO concedes as much, stating (p. 21) that a POSA would have understood PacBio-BAM as "indicat[ing] that those properties (IPD and PW) are preserved and available for use in base modification detection."

PO further states (pp. 22-23) that “PacBio-BAM suggests that the identity of the nucleotide should be considered in base modification detection.”

B. The Prior Art Does not Teach Away From “identity of the nucleotide”

PO repeatedly and incorrectly asserts (pp. 2, 13-15, 17-18, 20-21, 23-25, 28-29, 31) that the prior art, including the Flusberg Publications, “avoids” including identity of the nucleotide in input data structures. Moreover, PO’s arguments often appear to pretend claim 1 is limited to the specific data matrix and convolutional neural network (CNN) embodiment in Tse (EX1030), which is not the case.¹ The prior art teaches use of the claimed properties for detecting base modification and PO fails to address the full scope and content of the prior art.

i. The Flusberg Publications teach nucleotide identity

Contrary to PO’s argument (pp. 13-31), the Flusberg Publications broadly teach using combinations of reaction data, including nucleotide identity, obtained from SMRT sequencing to detect base modifications. *See Henny Penny Corp. v. Frymaster LLC*, 938 F.3d 1324, 1332 (Fed. Cir. 2019) (discussing “the

¹ Dr. Koo testified that inputting a data matrix having sequence information (nucleotide identity) into a model (CNN) was “traditional” and “widespread” in genomics prior to the ’794. EX1049, 134:6-135:3; *see also* EX1035, Fig. 1.

longstanding principle that the prior art must be considered for all its teachings, not selectively”).

Flusberg-338 directly teaches that a “combination of [its reaction data] can serve as a signal for the presence of a modification in the template nucleic acid.” EX1007, 14:10-14, 22:25-27. Flusberg-338 does not exclude nucleotide identity or position from this teaching.

Instead, it explains that

“[d]ifferent types of reaction data can be combined to detect

various kinds of modifications[.]”

EX1007, 14:53-62, 14:65-15:25

(“reaction data (e.g., sequence

reads, kinetics...”), Fig. 2 (reproduced herein). As shown in Fig. 2, the structure includes nucleotide identity and position, IPD, and PW. EX1007, 14:10-13

(“Reaction data includes data collected directly from the reaction, as well as the results of various manipulations of that directly collected data...”), 22:25-23:22;

see also EX1049, 103:15-106:17. As the petition established, the Flusberg

Publications disclose sequencing nucleic acid molecules using PacBio SMRT

sequencing, generating pulse traces in which each nucleotide produces a pulse

from which characteristic features may be derived. Pet., 27-28; EX1006, 462-63,

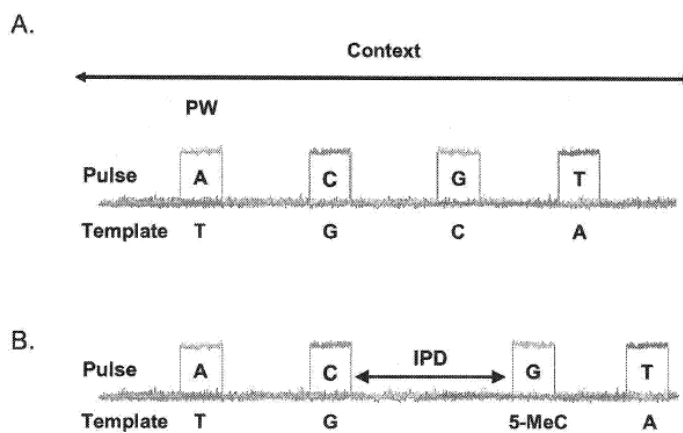


Figure 2

Fig. 1; EX1007, 23:14-22, 64:29-43, 66:46-67:4, Figs. 1-2; *see also* EX2037, 114:3-120:15 (explaining that pausing of the polymerase is indicative of base modification); EX1049, 141:20-25, 145:10-148:6. These pulse features include the identities of each nucleotide in the sequence, as well as PW and IPD for each nucleotide. EX1007, 64:29-43, 66:46-67:4, Figs. 1-2. Flusberg-338 even expressly describes inputting these pulse features, including nucleotide identity (basecalls), into the “algorithmic architecture” of a model (e.g., CRF) for base modification detection. EX1007, 64:29-43, 65: 4-16; 66:46-67:4; *see also* EX1037, 17-18, 36-39; EX1045, claim 1 (claiming inputting “basecalls; pulse widths; and interpulse durations” into a model for base modification detection). Accordingly, PO either ignores these teachings or provides incoherent distinctions, and is incorrect when it argues that the prior art fails to teach or suggest an input data structure comprising nucleotide identity.

Moreover, PO gives no good reason to ignore the scores of prior art teachings—including the Flusberg Publications—regarding the importance of sequence context, which includes both nucleotide identity and position. Dr. Koo attempts to brush aside such teachings, but none of his supposed justifications hold water. First, Dr. Koo makes a distinction of no moment in arguing sequence context is not limited to nucleotide identity. EX2041, ¶114-116 (“the term ‘local sequence context’...does not refer to nucleotide identity, but rather the window of

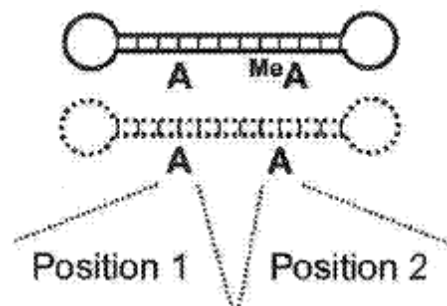
nucleotides around a modification”); POR, 29-31. The ’794 patent itself admits sequence context as providing both nucleotide identity and position. EX1001, 16:16-22; *see also* EX1007, Fig. 2; EX1049, 43:2-13, 106:8-17. Second, Dr. Koo argues that the prior art does not suggest to a POSA how “to use nucleotide identity.” EX2041, ¶114. Not only is this plainly wrong, but Element 1.2 does not require “use” of nucleotide identity; it only requires values for that property stored in an input data structure.

ii. Flusberg’s experimental example uses nucleotide identity

Much of PO’s argument seizes on a “specific example” using a known or fixed sequence, then leaps to the illogical conclusion that sequencing information is somehow “not used” in the Flusberg Publications and “avoided” entirely. *See, e.g.*, POR 17-18 (citing EX2041, ¶¶88, 107, 127, 132). PO’s blinkered argument is both wrong and ignores the scope of claim 1.

Flusberg-338 describes examples utilizing nucleotide sequence identity in methylation analysis. For example, Flusberg-338 exemplifies detecting 6-methyl adenine (6mA) modification in reaction data obtained from SMRT sequencing.

E.g., EX1007, 12:10-11, 12:46-50, 23:14-22, 65:33-63, 76:65-77:60, Figs. 2, 14. Flusberg-338 describes obtaining reaction data for two templates that are differentially methylated at an



A within the context GATC. EX1007, 77:3-6 (“The methylated template has an A within a GATC context at Position 1 and a ^mA within a GATC context at Position 2, whereas the unmethylated template has an A at both positions.”), Fig. 14 (reproduced in part herein). Flusberg-338 uses nucleotide identity when detecting 6mA at least by aligning the two template molecules as shown in Fig. 14. EX1049, 101:25-103:14. Flusberg-338 also uses nucleotide identity to verify the sequence context (e.g., GATC) for Position 1 and Position 2. See EX1007, Fig. 14; see also *id.*, Fig. 11; EX1006, Fig. 4.

When confronted during cross-examination, Dr. Koo could not and did not deny the use of nucleotide sequence identity. EX1049, 98:7-103:14. Instead, he dismissed it from consideration on the legally and factually erroneous basis that Fig. 14 was just a “toy schematic” and “not experimental data.” EX1049, 98:7-99:11. *Contra* EX1007, 65:46-52 (“FIG. 14 provides actual data....”). Even the examples in the Flusberg Publications use nucleotide identity and PO fails to demonstrate otherwise.

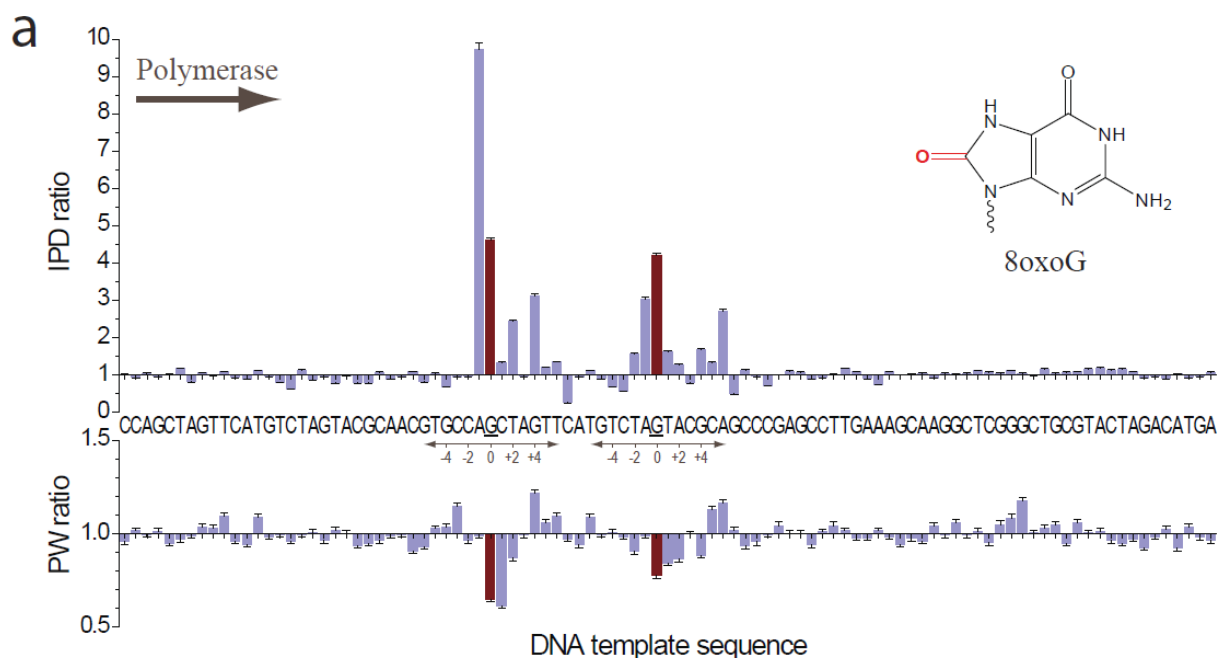
iii. Prior art corroborates obviousness

PO argues (pp. 3, 23) that prior art identified in the petition “uses nucleotide identity in the same way as the Flusberg publications.” Indeed, prior art echoes the Flusberg Publications in linking reaction data (e.g., sequence context, IPD, and PW) to base modifications and corroborates their teachings in this regard.

The kineticsTools manual describes using sequence context, including nucleotide identity, obtained from a BAM file to predict base modifications based on polymerase kinetics (IPD). The manual explains that the kineticsTools model “predicts the IPD using the **local sequence context** around the current position.” EX1021, 1. According to the manual, “kineticsTools inspects the alignment between the observed bases and the reference sequence—in order for an IPD measurement to be included in the analysis.” EX1021, 2. As Dr. Koo confirms, an alignment of observed bases to a reference sequence cannot be performed without knowing the nucleotide identity of the sequenced nucleic acid. EX1049, 85:20-89:5. The manual also explains that the model “includes a pre-trained lookup table mapping 12-mer DNA sequences to mean IPDs.” EX1021, 2. That is, kineticsTools takes 12-mer nucleotide sequences from the reaction data and compares it with identical sequences in the “lookup table.” Thus, kineticsTools plainly uses nucleotide identity, which it obtains from a BAM file, for base modification detection. *See id.*, 1-3; POR, 23 (acknowledging “kineticsTools receives the PacBio-BAM file as an input”); EX1049, 109:3-111:3.

Clark also describes using sequence context (among other properties) to identify various base modifications. Clark explains that the “magnitude and extent of the kinetic signature is dependent on the type of the base modification and the **local sequence context**.” EX1008, 2 (citing Flusberg-2010 as [5]). Clark describes

mapping reads obtained from SMRT sequencing to “the respective reference sequences.” *Id.*, 8. Clark further discusses analyzing pulses that were “aligned” to “the reference sequence.” *Id.* This is shown in Fig. 1 which depicts changes in pulse data (IPD and PW) for aligned modified and unmodified templates. EX1008, Fig. 1 (reproduced in part herein); *see* EX1049, 111:12-114:14.



Finally, the '794 patent itself contradicts PO's allegations regarding nucleotide identity. Specifically, the '794 patent acknowledges that Feng (EX1015) describes an input data structure including sequence context. EX1001, 19:2-11; *see also id.*, 16:16-22, 19:38-40 (describing the prior art as utilizing “sequence information in the neighboring nucleotides”), 19:14-50; EX1015, Title, Figs. 3-4; *Qualcomm Inc. v. Apple Inc.*, 24 F.4th 1367, 1376 (Fed. Cir. 2022) (“[A] patentee’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....”). Corroborating prior art and PO’s admission in the ’794 patents contradict PO’s litigation-driven assertion that the prior art “avoided” nucleotide identity.

C. The Prior Art Does not Teach Away From PW

PO argues (pp. 31-36) “the well-understood use of PW in base modification detection was to exclude it and instead include IPD.” PO’s argument fails at least because it is based on a myopic reading of the prior art ignoring the many teachings identifying PW for base modification detection. PO never addresses express disclosure in the Flusberg Publications describing using PW when detecting base 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, 35:13-16, 78:9-12 (“the presence of these modifications was shown to have an effect on pulse width.”), 78:40-42 (“8-oxoG altered pulse width over a window of 7-8 neighboring bases by as much as 40%.”), Figs. 2, 17, 19; *see also* EX1013, 194 (“Flusberg et al. used...IPD and PW associated with distinct modifications of deoxyribonucleotides.”). Even where presenting experimental examples using IPD, Flusberg-338 expressly states additional metrics are not excluded. EX1007, 67:64-67 (“it will be clear to one of ordinary skill that the approach can be extended to pulse metrics other than IPD, such as pulse width...etc.”). Nor does PO address

PacBio-BAM, which describes storing IPD and PW values in a manner compatible with “base modification detection.” EX1011, 8.

Additional PacBio publications presented in the petition further support the inclusion of PW in the data input structure. For example, Clark followed up on Flusberg’s observations and demonstrated that several DNA base modifications could be distinguished by their unique kinetic signatures. EX1008, Abstract, 7. For each modification, Clark assessed IPD and PW values in a window of nucleotides flanking the modified base. EX1008, Figs. 1-3.

PO’s argument boils down to cherry-picking a few embodiments where IPD was used then leaping to the illogical conclusion it was “well-established” to exclude PW. This is a non-sequitur and mischaracterizes the cited art. None of the prior art PO identifies states that PW should be excluded when detecting base modification. To the contrary, the PacBio White Paper, for example, discusses a “typical experiment” and cites bioinformatic tools utilizing, *inter alia*, PW, IPD, and sequence context. EX1016, 4 (citing EX1047); *see also* EX1047, 5 (“the most pertinent databases are: AlnArray, IPD..., and PulseWidth.”); EX1049, 114:22-128:13. Schadt expressly discloses that “reproducible variations in IPD and PW are induced as a result of modifications to the bases.” EX1010, 2 (citing Flusberg-2010). Schadt further describes the “strong dependence of IPD on local sequence context” and purports to model “sequence context” effects on IPD. *Id.* While

Schadt did not use PW in its “initial model,” *id.*, it by no means taught its exclusion.

The Flusberg Publications (among others) teach that both IPD and PW vary in a sequence context dependent manner and are all indicative of base modification. No more is needed to demonstrate obviousness.

III. NO OBJECTIVE INDICIA SUPPORTING NONOBVIOUSNESS

PO’s objective indicia arguments (pp. 36-65) mischaracterize evidence and fail at least due to a stunning lack of nexus with its exceptionally broad claims. EX1049, 149:9-22 (Dr. Koo did not even consider nexus); *Volvo Penta of the Ams., LLC v. Brunswick Corp.*, 81 F.4th 1202, 1211 (Fed. Cir. 2023) (no presumption where expert did not opine that product is coextensive with claims and patentee offered only one conclusory sentence).

First, PO argues (pp. 37-50) allegedly satisfying a previously unmet need for detecting 5mC (one type of modification) without bisulfite treatment. Claim 1 is not limited to 5mC (the only modification for which bisulfite mattered) and neither the claim nor the “input data structure” exclude bisulfite treatment prior to sequencing.

Next, PO argues (pp. 50-62) praise and copying of the Holistic Kinetic (HK) model in the Tse reference. But PO’s own expert admits that Tse describes one embodiment within the broad claim 1. EX2002, ¶60. Moreover, PO and its expert

both admit that Tse’s Keras CNN is an “open source” CNN that was already used in the prior art for identifying methylation sites and Dr. Wenger used a different “open source” CNN (TensorFlow). EX2041, ¶85. The only “praise” of record goes to PacBio’s own technology and aspects already in the prior art.

Finally, Patent Owner’s arguments (pp. 62-65) regarding alleged unexpected results are woefully deficient as they address neither the subject matter claimed nor “the closest prior art.” The argument disregards actual data from Flusberg-2010, ignores Flusberg-338 entirely, and draws limited comparison between a CNN and a principal component analysis modeling generated by the ’794 inventors themselves (inaccurately dubbed “Flusberg”). Moreover, there is no accuracy requirement in claim 1, which encompasses embodiments from the specification even less accurate than the “Flusberg” caricature.

It is well-established, “to be accorded substantial weight in the obviousness analysis, the evidence of secondary considerations must have a ‘nexus’ to the claims.” *Henny Penny*, 938 F.3d at 1332. As described in detail, PO here unquestionably fails to meet its burden of proving nexus and there is none. *Demaco Corp. v. F. Von Langsdorff Licensing, Ltd.*, 851 F.2d 1387, 1392 (Fed. Cir. 1988) (The burden of proof for nexus “resides with the patentee.”).

A. Bisulfite sequencing and modification

PO's argument (pp. 37-50) that it solved an unmet need for detecting of 5mC modifications without bisulfite conversion is remarkably disconnected from the scope of claim 1.

Initially, there is no dispute that bisulfite conversion is not relevant to modifications other than 5mC. EX1007, 2:28-29 (“[B]isulfite converts unmethylated cytosine, but not 5-MeC, to uracil...”); EX2039, 15 (“[B]isulfite-based sequencing cannot discriminate between 5mC and other modifications such as 5hmC, 5fC, and 5CaC.”); EX2038, 2 (“6mA remains undetectable with this approach.”); POR 38, 40 (acknowledging the same); EX1049, 43:23-48:11 (Dr. Koo acknowledging).

Yet, claim 1 is not limited to 5mC (or even methylation generally). Claim 1 broadly recites “detecting a modification of a nucleotide in a nucleic acid molecule.” The '794 patent describes “modifications” as including a wide variety of methylation and oxidation products that may be present in DNA or RNA. EX1001, 1:36-49, 2:44-48, 13:39-59, 14:12-14, 33:41-44, 109:20-26; EX1049, 25:12-26:12, 29:3-31:8. Claim 1 does not exclude bisulfite treatment and the specification expressly envisions treatment of samples prior to sequencing. EX1001, 2:39-44 (“[C]ertain pre-treatment...may serve to enhance aspects of the

invention....”), 87:21-24 (“[T]he first nucleic acid molecules may undergo a treatment such that the modification is present.”).

PO’s corresponding argument regarding alleged accuracy and lowered subread depth is not commensurate with claim 1 for additional reasons. POR 46 (“feature set”), 47 (“good correlation”), 48 (“lowered the subread depth”). Claim 1 is not limited to any particular “subread depth.” Furthermore, the input data structure “compris[es] a window of the nucleotides” not limited by size (1, 2, 3, 4, 5, etc.) or shape (upstream/downstream nucleotides). *E.g.*, EX1001, 27:39-51, 33:12-48; EX1049. 70:5-17, 72:12-73:7. PO touts the “predictive feature set” of Fig. 18 but ignores embodiments in the ’794 patent where different selected window sizes for the same “feature set” degraded predictive capacity to essentially a coin flip (i.e., AUC=0.50). *E.g.*, EX1001, Fig. 23; *see also id.*, Figs. 24, 25. With no accuracy requirement or restriction of window size, the input data structure of claim 1 includes these AUC=0.50 embodiments.

PO’s failure argument (pp. 42, 49-50) is head-scratching and unavailing. Direct detection of methylation without bisulfite sequencing is stated plainly in the title and abstract of Flusberg-2010. EX1006. During Dr. Dessimoz’s deposition, PO’s own lawyer even argued this as an “advantage” of both Flusberg publications. EX2037, 146:1-7. The PacBio White Paper corroborates, describing that “SMRT sequencing, in contrast [to second/next generation sequencing], does

not require base conversion within the source material in order to detect base modifications.” EX1006, 1. PO’s argument (pp. 48, 49) that the White Paper and kineticsTools discuss “250X minimum sequencing fold coverage per strand for 5mC” hardly demonstrates failure. It is also irrelevant given claim 1 is not limited to any depth of sequencing or to 5mC.

Beyond that, PO (pp. 42, 49) badly mischaracterizes Schadt. PO cites Schadt’s comment regarding “NGS technology” requiring bisulfite. But as is plainly apparent from the text and conceded by Dr. Koo during cross-examination, Schadt discusses “second (or next) generation (NGS) instruments” (e.g., Illumina) and then goes on to distinguish third generation PacBio sequencing. EX1010, 129; *see also id.*, Title (“Modeling kinetic rate variation in third generation DNA sequencing data...”), 137 (“here were applied to the only commercially available third-generation DNA sequencing instrument available at the time this manuscript was written (the PacBio RS)”); EX1049, 48:12-59:2. Furthermore, PO’s citation to Schadt’s comment regarding “no statistical framework” for base modification detection ignores the scope of claim 1 (e.g., encompassing a PCA “statistical framework”) and ignores Flusberg-338 entirely. Flusberg-338 plainly teaches use of statistical models, including a “conditional random field” or CRF model as in Schadt, as well as neural networks, for base modification detection. EX1007, 63:14-31, 64:28-58, 67:67-68:5.

B. Using a prior art model according to prior art teachings is obvious

PO's argument (pp. 50-62) regarding praise and copying of the Tse HK model fails as not commensurate with the scope of claim 1 and the unrebutted fact that PO identifies nothing not already found in the prior art.

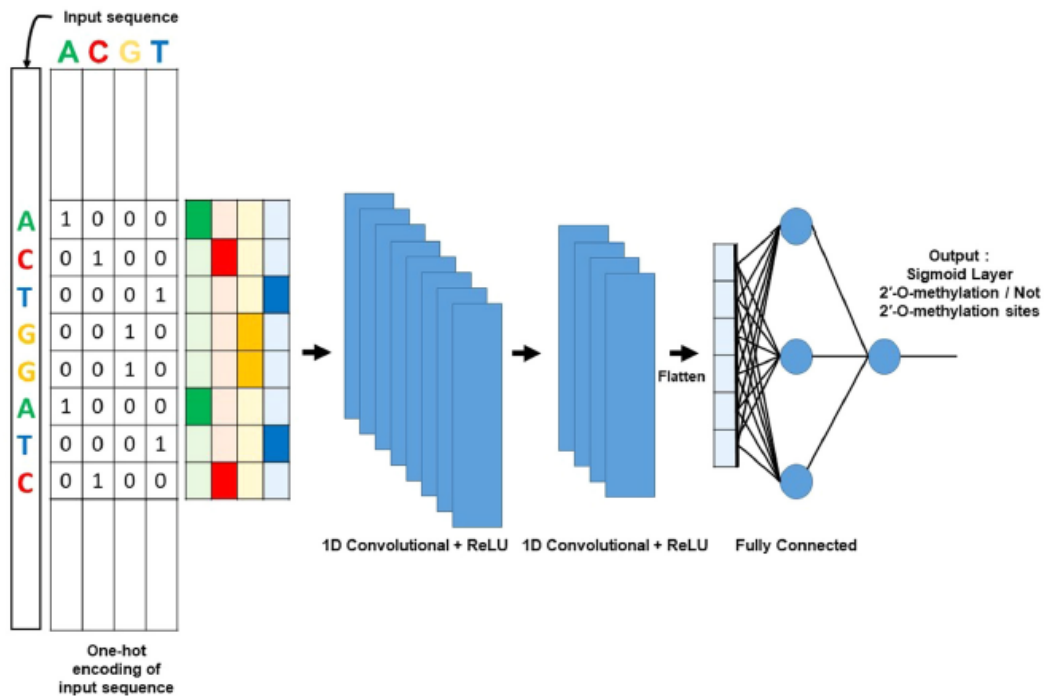
First, as Dr. Koo conceded in both his direct and deposition testimony, Tse's HK model is, at best, just one embodiment within the broad scope of claim 1. EX2002, ¶60 ("an embodiment of claim 1"); EX2002, 19:17-20:2 (testimony not withdrawn). In his second declaration, Dr. Koo repeatedly states claim 1 is not limited to an input matrix as in Tse. EX2041, ¶46 ("data matrix as an example"), ¶49 ("can be implemented in multiple ways"). Dr. Koo admitted that claim 1 is not limited to a CNN, let alone Tse's Keras CNN. EX1049, 76:25-78:12. Moreover, as discussed above, claim 1 is not limited to 5mC or any particular type of sequencing, and does not exclude bisulfite treatment of the "nucleotides sequenced." PO's arguments can be rejected for this reason alone, as no nexus exists. *Fox Factory, Inc. v. SRAM, LLC*, 944 F.3d 1366, 1377 (Fed. Cir. 2019) (coextensiveness requirement not met by arguing the patent claims broadly cover the product).

Second, PO is unable to point to any aspect of Tse's HK model that is not already found in the prior art. *See Fox Factory*, 944 F.3d at 1378 (secondary indicia must correspond to purportedly inventive aspect of combination, as opposed to unclaimed features or features found in prior art). As explained in the petition,

scores of prior art (including the Flusberg Publications) teach that sequence context, IPD, and PW are associated with modification status. Flusberg-338 specifically teaches using said features in various statistical models, including neural networks, for base modification detection. EX1007, 63:22-27 (“preferred algorithms include...neural networks”), 67:64-68:5 (“standard statistical classification algorithms, i.e....neural networks...”). Tse’s purported contribution was use of the Keras CNN in its modeling, but that too is admitted prior art. EX1030, 10 (“implemented on...Keras”); EX2041, ¶85 (Tse used “well known open source” Keras). Prior to the ’794 patent, the prior art already taught using Keras for identifying methylation sites from nucleotide sequence. EX1035, 2, 3 (“[T]he proposed model is implemented by Keras framework....”); EX1049, 129:8-130:11. “Objective indicia” cannot save PO’s claims where it does no more than follow the teachings of the prior art.

PO curiously alleges copying of Tse’s HK model by Dr. Wenger but readily admits he used a different data structure (e.g., 16 versus 21 nucleotide window, different matrix organization) and a different CNN (“TensorFlow CNN”). EX2041, ¶¶81-83, 85. PO attempts to brush over the readily apparent lack of copying by alleging “striking similarity” between Tse’s HK model and Dr. Wenger’s model. POR, 55. For example, PO (pp. 56-57) identifies a matrix “input layer,” “convolutional layers,” and “sigmoid function” as common to the different models.

First, none of these aspects are recited in the claims. *See Windsurfing Int’l, Inc. v. AMF, Inc.*, 782 F.2d 995, 1000 (Fed. Cir. 1986) (“copying the **claimed invention**”). Additionally, Dr. Koo admits these “striking similarities” as conventional structure for any prior art CNN. Indeed, the Keras model was already implemented by Tahir for methylation identification and similarly illustrates a matrix “input layer,” “convolutional layers,” and an “output layer based on a sigmoid function.” EX1035, Fig. 1 (below); EX1049, 81:18-83:6, 129:3-135:23. Dr. Wenger used a prior art “TensorFlow CNN” that, unsurprisingly, reflects these conventional CNN structures.



PO alleges praise but fails to demonstrate that this praise relates to any inventive aspects of Tse or claim 1 (there are none). For example, PO argues that

PacBio credited Dr. Lo's lab with "development of algorithms that leverage HiFi sequencing information from both DNA strands to improve the detection of 5-methyl-c." EX2023. But PO's claims do not require HiFi sequencing, 5mC detection, or information from both DNA strands. Nor does claim 1 exclude bisulfite sequencing. As for the algorithms Tse used in the open-source Keras CNN, those are both unclaimed and in the prior art.

PO attempts to misappropriate "praise" directed to PacBio's own technology. PO cites a 2020 earnings call where Mr. Henry discussed "the power and the potential of PacBio HiFi² sequencing technology." EX2017, 5. Mr. Henry praised the newly released HiFi sequencing on Sequel II device, which allows for "comprehensive human variant detection" and epigenetic analyses. EX1048 ("eight times the scale"). In this context, Mr. Henry shared a "customer highlight" where Dr. Lo (a PacBio customer) used the PacBio HiFi sequencing product following

² HiFi sequencing is a PacBio sequencing product. Concurrent with the 2019 Sequel II release, PacBio provided improved HiFi base calling to customers using "DeepVariant," a deep learning algorithm. EX1032, 2, 7. Dr. Koo testified he was unfamiliar with PacBio's HiFi sequencing, its release prior to the '794 patent, and that EX1032 (like many other petition-cited documents) was withheld from his consideration. EX1049, 36:4-23, 140:5-9.

the teachings of PacBio publications like those cited in the petition. Mr. Henry also congratulated another customer at HudsonAlpha Institute of Biology “for their work in leveraging PacBio Hifi.” EX2017, 5. The identified “praise” is attributed to PacBio’s own business and products, and PO fails to demonstrate otherwise.

C. No unexpected results

PO argues (pp. 62-65) unexpected results but addresses neither the scope of claim 1 nor the closest prior art as required.

PO points to Fig. 29 and argues the CNN model is “far better” than a PCA model, but its claims are not limited to a CNN model and even cover a PCA model. EX1001, claim 11 (“a principal component analysis”). Claim 1 has no accuracy limitation, and PO ignores other ’794 patent embodiments that are within claim 1 but lack the touted accuracy. *See* EX1001, 36:5-29, Fig. 30 (AUC=0.83 for logistic regression); *see also id.*, claim 11 (“the model includes a...logistic regression”), Fig. 23 (AUC=0.50). PO’s argument is disconnected from claim 1 and fails to establish unexpected results for this reason alone. *Cf. Millennium Pharms., Inc. v. Sandoz Inc.*, 862 F.3d 1356, 1368 (Fed. Cir. 2017) (“Nonobviousness may be established when ***an invention*** yielded more than predictable results.” (cleaned up)).

As to the prior art, PO argues (p. 63) the Flusberg Publications are the closest prior art but fails to address either reference. PO refers instead to data

reported in the '794 patent (Fig. 29). To be clear, Fig. 29 is not data reported in either Flusberg Publication but instead was generated by inventors of '794 patent—individuals with a vested interest in the outcome of this proceeding. There is no description of this inventor-generated experiment that complies with 37 C.F.R. §42.65 or would even allow a POSA to scrutinize whether any aspect of Flusberg-2010 was faithfully followed. The '794 patent does not disclose read length or depth of sequencing, for example, whereas Flusberg-2010 discloses that increased sequencing allows better separation of methylations. *Compare* EX1001, 35:21-37, *with* EX1006, 463 (“As we collected more circular subreads....[t]his substantially improved discrimination....”), 464 (“[L]onger read lengths would yield even better discrimination.”). In any event, Flusberg-2010 reports experiments distinguishing both 5mC and mA, prior art content relevant to claim 1 (“modification”) that does not disappear based on the inventors self-interested reporting. EX1006, Figs. 3-4; *see id.*, 463-64 (0.96 AUC for mA detection). The cited Fig. 29 is not “Flusberg,” and this cryptically described inventor-generated figure deserves no weight. *In re Michalek*, 162 F.2d 229, 232 (C.C.P.A. 1947) (“[T]he failures of experimenters who have no interest in succeeding should not be accorded great weight....”). Moreover, the Flusberg Publications are not limited to PCA modeling (which is encompassed by claim 1 anyway). As discussed above, Flusberg-338 describes use of various “standard statistical classification

algorithms” including neural networks. EX1007, 67:67-68:5. PO does not compare any results to the closest prior art.

Accordingly, PO fails to establish any of its misguided secondary considerations arguments have nexus with claim 1. Claims 1-19 are obvious and PO fails to demonstrate otherwise.

IV. DEPENDENT CLAIMS

PO advances no separate argument regarding dependent claims 2-19. In fact, Dr. Koo acknowledged he did not even consider references cited in the petition regarding dependent claims. EX1049, 139:23-140:23, 150:7-19. As such, the argument in the petition regarding dependent claims 2-19 is unaddressed and un rebutted.

V. CONCLUSION

Claims 1-19 of the '794 patent are unpatentable and should be canceled.

Respectfully submitted,

Dated: October 28, 2024

/ Michael T. Rosato /

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

VI. CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. §42.24(d), the undersigned certifies that this Reply complies with the type-volume limitation of 37 C.F.R. §42.24(c). The word count application of the word processing program used to prepare this Reply indicates that the Reply contains 5,560 words, excluding the parts of the brief exempted by 37 C.F.R. §42.24(c).

Respectfully submitted,

Dated: October 28, 2024

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

VII. APPENDIX – PETITIONER’S UPDATED EXHIBIT LIST

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/Pacific-Biosciences/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 (Jan 25, 2017)
1043	GitHub, PacBio BAM 5.0, https://github.com/PacificBiosciences/PacBioFileFormats/blob/5.0/BAM.rst (Jan 25, 2017)
1044	PacBio Sequencing – How it Works video file
1045	U.S. Patent No. 11,844,666 to Flusberg et al. (“Flusberg-666”)
1046	U.S. Patent Application No. 17/891,899 File History
1047	DNA Modification Detection with SMRT® Sequencing Using R (November 15, 2011), (archived Aug. 13, 2017, at http://web.archive.org/web/20170813094741if_/https://raw.githubusercontent.com/PacificBiosciences/R-kinetics/master/analysis.pdf)
1048	Pacific Biosciences Press Release – New Sequel II System, Featuring ~8 Times the DNA Sequencing Data Output, https://www.pacb.com/press_releases/pacific-biosciences-launches-

	<u>new-sequel-ii-system-featuring-8-times-the-dna-sequencing-data-output/</u> (April 24, 2019)
1049	Deposition Transcript of Dr. Peter Koo (Oct. 15, 2024)

CERTIFICATE OF SERVICE

The undersigned certifies that the foregoing Petitioner's Reply to Patent Owner's Response and Exhibits 1047-1049 were served on October 28, 2024, at the following electronic service addresses:

Michael J. Wise
Joseph P. Hamilton
Yingli Wang
Perkins Coie LLP
wise-ptab@perkinscoie.com
hamilton-ptab@perkinscoie.com
wang-ptab@perkinscoie.com

Respectfully submitted,

Dated: October 28, 2024

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