

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

PTAB Case No. IPR2024-00028
U.S. Patent No. 11,091,794

PATENT OWNER'S PRELIMINARY RESPONSE

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. THE '794 PATENT	4
A. Claim 1 recites a data structure for training a model that has the same structure as the “input data structure.”	8
B. The '794 patent examples disclose a data structure for training a model that has the same structure as the “input data structure.”	9
C. Claim Construction.....	10
D. Person of Ordinary Skill in the Art	10
III. THE BOARD HAS DISCRETION TO DENY INSTITUTION PURSUANT TO §325(D)	11
A. The Petition presents substantially the same art and arguments that were previously presented to the Office.....	14
B. The Petition has not demonstrated that the Office erred in a manner material to the patentability of the challenged claims.....	22
IV. CLAIM 1 IS NOT OBVIOUS OVER GROUND 1.....	26
A. PacBio-BAM does not disclose or render obvious the “input data structure” or its combination of features recited in claim 1.	30
1. The PacBio-BAM contains extraneous information that would render training a model inoperable.	31
2. The PacBio-BAM does not include a “target position.”	32
3. The PacBio-BAM cannot be used to train a model.	33
B. Flusberg and the '338 patent do not disclose or render obvious the “input data structure” or its combination of features recited in claim 1.	35

TABLE OF CONTENTS (CONTINUED)

	Page
1. Flusberg does not disclose an “input data structure” including the “identity of a nucleotide” and teaches away from including pulse width.	35
2. The '338 patent does not disclose an “input data structure” including the “identity of a nucleotide” and teaches away from including pulse width.	39
C. Petitioner provides no motivation to combine Flusberg in the three-reference combination of Ground 1.	45
D. Modification detection using the claimed input data structure provided unexpected results with higher accuracy compared to the cited prior art.	47
E. Objective indicia strongly favors nonobviousness of claim 1.	48
1. Petitioner’s long felt but unsolved need	50
2. Petitioner’s praise for the claimed invention.	53
3. Petitioner copied the claimed invention as set forth in Tse.	55
4. Nexus between Tse and claim 1.	56
V. CONCLUSION.....	60

EXHIBIT LIST

	Petitioner's Exhibits
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")

Petitioner's Exhibits	
No.	Description
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”)

Petitioner's Exhibits	
No.	Description
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")

	Petitioner's Exhibits
No.	Description
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

Patent Owner's Exhibits	
No.	Description
2001	Color Drawings of US 11,091,794 Downloaded from USPTO Patent Center
2002	Declaration of Dr. Peter Koo
2003	Dr. Peter Koo's CV
2004	Li et al. and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map format and SAMtools," <i>Bioinformatics</i> 25(16): 2078-2079 ("Li")
2005	Suzuki et al., "AgIn: measuring the landscape of CpG methylation of individual repetitive elements," <i>Bioinformatics</i> , 32(19), 2016, 2911–2919 ("Suzuki")
2006	Zheng and Casari, "Feature Engineering for Machine Learning," First Edition, O'Reilly Media, Inc., Sebastopol, CA, 2018 (Excerpts)
2007	Web page (print friendly screen capture): Johns Hopkins University Center for Computational Biology, StringTie – Transcript assembly and quantification for RNA-Seq. URL: http://ccb.jhu.edu/software/stringtie/index.shtml . Web page visited January 23, 2024
2008	Web page (print friendly screen capture): Broad Institute Genome Analysis Toolkit. URL: http://gatk.broadinstitute.org/hc/en-us . Web page visited January 23, 2024
2009	Web page (print friendly screen capture): Broad Institute Genome Analysis Toolkit - HaplotypeCaller. March 4, 2020. URL: http://gatk.broadinstitute.org/hc/en-us/articles/360040096812-HaplotypeCaller . Web page visited January 23, 2024
2010	Web page: University of California at Santa Cruz, Genome Browser BAM Track Format. URL: https://genome.ucsc.edu/goldenPath/help/bam.html . Web page visited January 23, 2024

Patent Owner's Exhibits	
No.	Description
2011	Web page (print friendly screen capture): Git Hub HPC Training – Peak calling with MACS2. June 12, 2017. URL: https://hbctraining.github.io/Intro-to-ChIPseq/lessons/05_peak_calling_mac.html . Web page visited January 23, 2024
2012	YouTube Video: PacBio Data Deep Dive: A Closer Look at HiFi Sequencing. PacBio May 24, 2021. URL: https://www.youtube.com/watch?v=1aqnLQ-nwYk . Downloaded January 22, 2024. (Also available at https://www.pacb.com/videos/pacbio-data-deep-dive-a-closer-look-at-hifi-sequencing/)
2013	Screen Capture of Exhibit 2012 (YouTube) https://www.youtube.com/watch?v=1aqnLQ-nwYk
2014	Screen Capture of Exhibit 2012 (PacBio.com) https://www.pacb.com/videos/pacbio-data-deep-dive-a-closer-look-at-hifi-sequencing/
2015	Transcript of 9:50 – 11:45 of Exhibit 2012
2016	<i>Take2 Technologies Limited v. Pacific Biosciences of California, Inc.</i> , U.S. District Court for the District of Delaware, Case No. 5:23-cv-04166-EJD, Dkt. 8, Complaint (Public Redacted Version), December 14, 2022
2017	<i>Take2 Technologies Limited v. Pacific Biosciences of California, Inc.</i> , 5:23-cv-04166-EJD, Dkt. 8-2, Exhibit C to the Complaint – Transcript of PacBio Q4 2020 Earnings call on February 10, 2021 (available at https://www.fool.com/earnings/call-transcripts/2021/02/10/pacific-biosciences-of-californiapacb-q4-2020-ear/) (Complaint Ex. C)

Patent Owner's Exhibits	
No.	Description
2018	<i>Take2 Technologies Limited v. Pacific Biosciences of California, Inc.</i> , 5:23-cv-04166-EJD, Dkt. 8-2, Exhibit E to the Complaint –PacBio Children's Mercy Kansas City Press Release dated April 21, 2022 (available at https://www.pacb.com/press_releases/pacbio-and-childrens-mercy-kansas-city-expand-collaboration-taking-a-multi-omics-approach-to-characterize-rare-disease/) (Complaint Ex. E)
2019	<i>Take2 Technologies Limited v. Pacific Biosciences of California, Inc.</i> , 5:23-cv-04166-EJD, Dkt. 8-2, Exhibit F to the Complaint - Screenshot PacBio Products - Epigenetics (available at https://www.pacb.com/products-and-services/applications/epigenetics/) (Complaint Ex. F)
2020	YouTube Video: "HiFi Sequencing: See What You've Been Missing. PacBio, November 2, 2021. URL: https://www.youtube.com/watch?v=d2an7LXhPSU . Downloaded January 19, 2024. (Also available at https://www.pacb.com/videos/hifi-sequencing-see-what-youve-been-missing/)
2021	Screen Capture of Exhibit 2020 (YouTube) https://www.youtube.com/watch?v=d2an7LXhPSU
2022	Screen Capture of Exhibit 2020 (PacBio.com) https://www.pacb.com/videos/hifi-sequencing-see-what-youve-been-missing/
2023	Transcript of 05:05-06:06 from Exhibit 2020
2024	YouTube Video: "Methylation Detection with PacBio HiFi Sequencing: ASHG 2021" PacBio, October 27, 2021. URL: https://www.youtube.com/watch?v=rT23lBCp6Ro . Downloaded January 17, 2024.
2025	Screen Capture of Exhibit 2024 (YouTube) https://www.youtube.com/watch?v=rT23lBCp6Ro

Patent Owner's Exhibits	
No.	Description
2026	Transcript of 15:46–15:56 from Exhibit 2020
2027	Web page: PacBio Sequencing 101: The SMRT Cell in PacBio long-read sequencing. URL: https://www.pacb.com/blog/smrt-cell/ . Web page visited January 23, 2024
2028	<i>Take2 Technologies Limited v. Pacific Biosciences of California, Inc.</i> , 5:23-cv-04166-EJD, Dkt. 8-2, Exhibit D to the Complaint – PacBio SMRT® Link release notes (v11.0) (Complaint Ex. D)
2029	U.S. Patent Application Publication US2010/0221716 by Flusberg
2030	Declaration of Amy Candeloro
2031	Declaration of Chelsea Baker
2032	List of Materials Considered by Dr. Peter K. Koo

I. INTRODUCTION

Patent Owner requests that the Petition be denied (1) pursuant to 35 U.S.C. § 325(d) because the Petition relies on the same art and arguments for Ground 1 previously considered by the Office and fails to demonstrate that the Office erred in a way material to patentability and (2) under 35 U.S.C. § 314(a) because the Petition fails to establish a reasonable likelihood to prevail on Ground 1 with respect to claim 1, the sole independent claim in the '794 patent. Because all other asserted Grounds rely upon prevailing on Ground 1 for claim 1, all other grounds fail for the same reasons.

Claim 1 requires, in part, creating an input data structure having a specific combination of features as follows:

(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 [PW] corresponding to the nucleotide, and

the interpulse duration [IPD];

EX1001, claim 1.

Ground 1 includes three of Petitioner's own references: Flusberg, the '338 patent and the PacBio-BAM. During prosecution, the Office extensively

considered two of those references, Flusberg and the '338 patent, and allowed claim 1 over those references. Specifically, the examiner recognized that despite using the data generated in the normal course of SMRT sequencing (EX1004, 803, 806), Flusberg and the '338 patent do not appear to “measure the widths as a data point to be combined with interpulse duration, nucleotide identity and position in determining the presence of a mutation.” EX1004, 217. In other words, neither reference discloses the combination of features of the “input data structure” in claim 1.

Here, Petitioner requests that the Board reconsider the Office’s determination of patentability based upon purportedly new information, the PacBio-BAM. But, contrary to Petitioner’s assertion, Patent Owner’s infringement contentions do not allege that the claimed “input data structure” is the PacBio-BAM. Instead, the contentions allege that the input data structure is created from the data in the PacBio-BAM. EX1014, 25 (“The HiFi reads in the BAM file are then used to create the data structure of the claimed element [1.b], i.e., an input data structure, for claim element [1.c].”).

And, as explained by Petitioner’s own expert, the PacBio-BAM is “the standard data structure in which PacBio SMRT sequencers store data” (EX1002, ¶ 81) that is “generated in the normal course of SMRT sequencing” (EX1002, ¶ 52). As such, the PacBio-BAM is cumulative to Flusberg and the '338 patent, which

both disclose using the data generated in the normal course of SMRT sequencing. So, Petitioner's purportedly new disclosure, the PacBio-BAM, collapses to the same issue already considered and rejected by the Office, namely, despite having the data generated in the normal course of SMRT sequencing, claim 1 is patentable over Flusberg and the '338 patent.

In fact, later publications from some of the same coauthors of both Flusberg and the '338 patent confirm that Ground 1 does not render the combination of features in claim 1 obvious. In fact, three years after Flusberg and the '338 patent, coauthors from both references explained that as of that date, "no statistical framework has been proposed to identify kinetic variation events from SMRT sequencing data that can serve as a surrogate for chemical modifications of DNA or RNA sequences." EX1010, 130. So, those coauthors explain that "the modeling presented herein [again, three years after Flusberg and the '338 patent,] should be considered as a first step in leveraging the extensive kinetic information provided by SMRT sequencing...." *Id.*, 138. That first step, however, does not include the claimed combination of features that Petitioner alleges is obvious over Flusberg and the '338 patent, and instead, the coauthors "restrict attention to IPDs [interpulse duration]." *Id.*, 130.

Finally, objective indicia from Petitioner strongly favors nonobvious. For example, following publication of the invention, Petitioner's own president and

CEO praised the inventors of the '794 patent for their “development of a new method for highly accurate simultaneous determination of DNA sequence and CPG methylation in one go[]” that will “help researchers explore the impact of epigenetic changes in humans and other organisms and has great potential in clinical research and eventually diagnostics in cancer and other disease areas where methylation changes are known to be important disease markers.” EX2017, 5.

For the forgoing reasons Patent Owner respectfully requests that the Petition be denied.

II. THE '794 PATENT

The '794 patent discloses, *inter alia*, methods for determining nucleotide modifications, *e.g.*, methylation of DNA nucleotides. Methylation of DNA nucleotides is essential for mammalian development and is involved in a variety of biological functions and diseases. EX1001, 1:61-2:2.

Claim 1 of the '794 patent relates to the discovery of a method employing a unique combination of some of the data from SMRT sequencing that could be used to detect base modifications, and in particular detect methylated cytosine, to a much higher degree of accuracy than prior art methods. In claim 1, a unique combination of data is recited in an “input data structure,” which includes a window of nucleotides around a target nucleotide (the putative modified base) and for each nucleotide in that window the identity of the nucleotide (A, C, G or T), the

position of the nucleotide relative to the target, the width of the pulse associated with the nucleotide (PW), and the interpulse duration (IPD). EX2002, ¶25. The measurement window of Figure 4 of the '794 patent illustrates an exemplary input data structure:

Measurement window 400

	G		A		T		N		A		C		T		402
	-3		-2		-1		0		+1		+2		+3		404
	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW	408
A	0	0	1.0	0.3	0	0	0	0	1.1	0.8	0	0	0	0	412
C	0	0	0	0	0	0	0	0	0	0	2.3	0.8	0	0	416
G	1.5	1.7	0	0	0	0	0.9	1.2	0	0	0	0	0	0	420
T	0	0	0	0	1.2	0.2	0	0	0	0	0	0	1.5	1.7	

EX2001, FIG. 4.¹

The top row of the array identifies the target (N) and the detected nucleotides, the second row identifies the relative position (+1, +2, etc.) to the target (0). Each position has two columns that include the pulse width (PW) and interpulse duration (IPD) values for that position recorded in rows (408, 412, 416, and 420). EX1001, 23:28-51.

¹ Citations to the '794 patent are to EX1001, except, citations to the '794 patent figures are to EX2001, which is a color copy of the '794 patent figures.

Using over 800,000 methylated and unmethylated DNA sequences to train a model, the inventors investigated the predictive power of the features and feature combinations. EX2002, ¶47. “We evaluated the power of each feature (sequence context, IPD and PW) in predicting the methylation state of C...” EX1001, 32: 18-20.

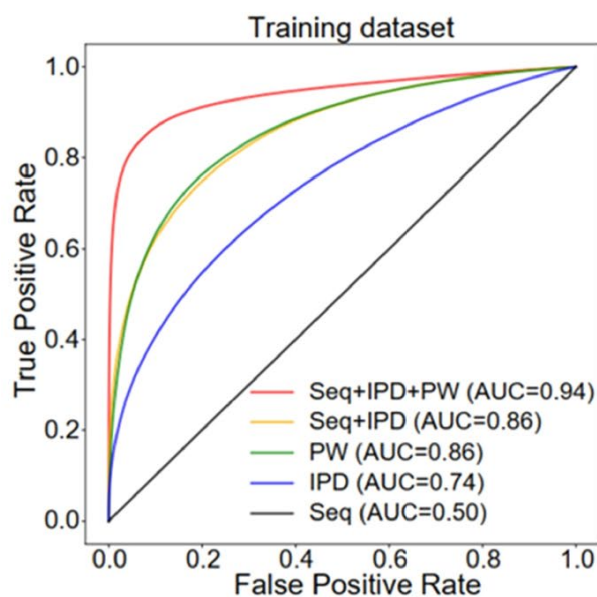


FIG. 18A

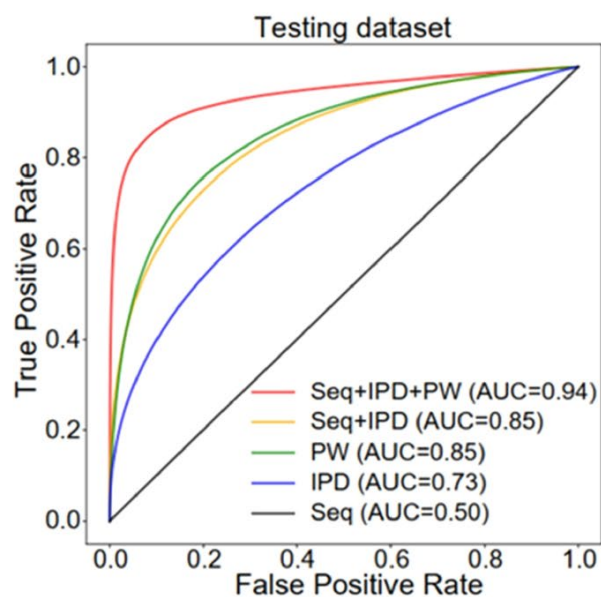


FIG. 18B

EX2001, FIGs. 18A, B.

As shown in Figure 18, identity of the nucleotide and relative position to the target (together “Seq”), IPD and PW (Seq+IPD+PW) achieved the highest accuracy (AUC = 0.94) over the identity of the nucleotide and relative position to the target (together “Seq”) and IPD (Seq+IPD), and PW, IPD, and identity of the nucleotide and relative position to the target (together “Seq”), respectively.

Accordingly, FIG. 18 shows the strong predictive power of claim 1 to detect a modification. EX2002, ¶¶49, 143. In the '794 patent “[s]equence context refers to the base compositions (A, C, G or T) and the base orders in a stretch of DNA.” EX1001, 19:60-63. In other words, the '794 patent uses the term “sequence context” to refer to both the identity of the nucleotide and the position of the nucleotide relative to a target. EX2022, ¶31.

The inventors investigated the predictive power of other features including subread depth, use of Watson and/or Crick strands, window sizes and window locations. EX2002, ¶¶50-56. The inventors systematically probed the predictive power of a variety of features to construct an optimal input data structure for a trainable model that accurately detects base modification. EX2002, ¶¶35, 48. But the inventors did not stop there. The inventors used the trained model with the claimed input data structure to detect modifications in a variety of practical applications including imprint regions in placental tissue (44:50-47:35), maternal circulating cell free DNA (47:36-48:63), copy number changes in tumor tissue (49:56-51:14), maternal plasma DNA tissue mapping (51:15-52:32), regional methylation quantification (52:33-60:17), differential methylation between tumor normal tissue (60:18-61:60), and viral DNA methylation (61:61-62:48). EX1001, EX2002 ¶¶56. In short, the '794 patent is a tour de force of base modification detection.

A. Claim 1 recites a data structure for training a model that has the same structure as the “input data structure.”

Claim 1 recites “a method for detecting a modification of a nucleotide in a nucleic acid molecule” including creating an “input data structure” that comprises “a window of nucleotides” surrounding a target position in the nucleic acid (“a target position within the window”). EX1001, claim 1(b). Step (d) recites determining ... whether the modification is present in a nucleotide at the target position...” For each nucleotide in the window, the “input data structure” includes “the identity of the nucleotide,” the “position of the nucleotide with respect to a target position within the window,” the “width of the pulse corresponding to the nucleotide (PW)” and “the interpulse duration (IPD)”. EX1001, claim 1(b). EX2002, ¶36.

Claim 1 further requires a model that is “trained by” “receiving a plurality of first data structures” which have the same structure and properties as the “input data structure” set forth in claim 1(b) as the claim recites “each first data structure comprising values for the same properties as the input data structure.” These data structures are used to train the model as the claim explains “storing a plurality of first training samples, each including one of the first plurality of first data structures....” and “optimizing, using the plurality of first training samples, parameters of the model...” EX2002, ¶37. Both experts agree that a model should be trained by data structures that match the input data structure used for

modification detection. EX2002, ¶38; EX1002 ¶130.

B. The '794 patent examples disclose a data structure for training a model that has the same structure as the “input data structure.”

The '794 patent discloses that “[k]inetic features may be obtained for a measurement window of a certain number of nucleotides upstream and downstream of the target bases. These features (e.g., at particular locations in the measurement window) can be used to train a machine learning model.” EX1001, 22:65-23:4.

As illustrated in FIG. 4, the '794 patent discloses an input data structure of a window of nucleotides (“measurement window 400”), a target position (position 0, nucleotide identity N), the identity of each nucleotide (row 402), the nucleotide position relative to target (row 404), the IPD and PW values (rows 408, 412, 416, and 420). EX1001, 23:28-51, EX2002, ¶¶40-43.

Measurement window 400

	G		A		T		N		A		C		T	
	-3		-2		-1		0		+1		+2		+3	
	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW
A	0	0	1.0	0.3	0	0	0	0	1.1	0.8	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	2.3	0.8	0	0
G	1.5	1.7	0	0	0	0	0.9	1.2	0	0	0	0	0	0
T	0	0	0	0	1.2	0.2	0	0	0	0	0	0	1.5	1.7

“[A]t the relative position of 0, the IPD and PW values were shown in the

row indicating ‘G’ ... suggesting that a guanine was called in the sequence result at that position. The other grids in a column that did not correspond to a sequence base would be coded ‘0.’” EX1001, 23:28-51.

The ’794 patent teaches additional input data structures for training a model created using sequencing data from the Watson strands (FIG. 4), the Crick strand (FIG. 5), or both Crick and Watson strands (FIGs. 6-8). EX1001, 23:14 to 24:56, EX2002, ¶43. The ’794 patent describes and illustrates how to train a model and use the trained model to determine a modification in a nucleic acid using the same input data structures. EX1001, FIGs. 10, 12, 14 and 16 (testing) and FIG. 9, 11, 13, and 15 (training). EX1001, 25:57-30:59, EX2002, ¶44-48.

C. Claim Construction

In an *inter partes* review proceeding, claims are construed using the same claim construction standard used in a civil action under 35 U.S.C. 282(b), including construing the claim in accordance with the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent. 37 CFR § 42.100(b). That standard is applied herein.

D. Person of Ordinary Skill in the Art

For the purposes of this response, Patent Owner has applied Petitioner’s definition of a person of ordinary skill in the art (“POSITA”). Petition, 26.

EX2002, ¶22.

III. THE BOARD HAS DISCRETION TO DENY INSTITUTION PURSUANT TO §325(D)

Institution denial under §325(d) is warranted when (1) the Petition relies upon the same or substantially the same art or the same or substantially the same arguments previously presented to the Office; and (2) the Petition fails to demonstrate that the Office erred in a manner material to the patentability of challenged claims. *Advanced Bionics, LLC v. Med-El Elektromedizinische Geräte GmbH*, 2020 WL 740292 at *3 (PTAB 2020) (precedential).

Within this two-part framework, the Board considers a number of non-exclusive factors in evaluating whether to exercise its discretion under § 325(d). *See Becton, Dickinson & Co. v. B. Braun Melsungen AG*, 2017 WL 6405100 *6 (PTAB 2017). The *Becton, Dickinson* factors are as follows:

- a. the similarities and material differences between the asserted art and the prior art involved during examination;
- b. the cumulative nature of the asserted art and the prior art evaluated during examination;
- c. the extent to which the asserted art was evaluated during examination, including whether the prior art was the basis for rejection;
- d. the extent of the overlap between the arguments made during examination and the manner in which petitioner relies on the prior art or patent owner distinguishes the prior art;

e. whether petitioner has pointed out sufficiently how the examiner erred in its evaluation of the asserted prior art; and

f. the extent to which additional evidence and facts presented in the petition warrant reconsideration of the prior art or arguments.

Becton, Dickinson, 2017 WL 6405100 at *6. Factors (a), (b), and (d) relate to whether the art or arguments presented in the Petition are the same or substantially the same as those previously presented to the Office. *Advanced Bionics*, 2020 WL 740292 at *4. And, factors (c), (e), and (f) “relate to whether the petitioner has demonstrated a material error by the Office” in its prior consideration of that art or arguments. *Id.*

Here, Ground 1 is the sole proposed ground for independent claim 1, from which all other claims depend. Because Ground 1 relies upon the same art and arguments previously presented to the Office, Patent Owner requests that institution be denied. During prosecution the Examiner allowed claim 1 over Flusberg and the '338 patent, finding that while “Flusberg displays various widths of pulses in their sequence data in a window,” Flusberg does not appear to “measure the widths as a data point to be combined with interpulse duration, nucleotide identity and position in determining the presence of a mutation.” EX1004, 217. The examiner references that same conclusion with respect the '338 patent. *Id.* In other words, while Flusberg and the '338 patent both disclose

receiving the standard output of SMRT sequencing, claim 1 is patentable over those references because the references do not disclose using that standard output of SMRT sequencing to create an “input data structure” having the claimed combination of features including pulse width, interpulse duration, nucleotide identify and position to determine the presence of a mutation.

Because Petitioner’s same arguments and references were previously overcome during prosecution, Petitioner attempts to manufacture a new issue and reference, namely, that the Office did not consider the PacBio-BAM. But as Petitioner’s own expert explains, the PacBio-BAM is “the standard data structure in which PacBio SMRT sequencers store data.” EX1002, ¶81. That data is “generated in the normal course of SMRT sequencing.” EX1002, ¶52. As such the PacBio-BAM is cumulative to the disclosures of Flusberg and the ’338 patent and adds nothing to the question of whether those references render obvious the claimed “input data structure” having its claimed combination of features.

EX2002, ¶31

Petitioner’s argument for why the PacBio-BAM is purportedly new is premised upon Petitioner’s misrepresentation of Patent Owner’s infringement contentions in a companion litigation. Petition, 22. In short, Petitioner claims that Patent Owner admitted that the standard output of PacBio SMRT sequencing, the PacBio-BAM, is the claimed “input data structure” having the claimed

combination of features. Nonsense. Patent Owner’s infringement contentions expressly recite that the PacBio-BAM is “used to create the data structure of the claimed element [1.b], i.e., an input data structure, for claim element [1.c].”

EX1014, 25. The contentions do not alleged that the PacBio-BAM is the claimed input data structure.

As set forth below, application of the *Becton, Dickson* factors favors denying the Petition.

A. The Petition presents substantially the same art and arguments that were previously presented to the Office.

First, *Becton, Dickinson* factor a—the similarities and material differences between the asserted art and the prior art involved during examination—favors denial. The two primary references of Ground 1, Flusberg and the ’338 patent, were expressly relied upon by the examiner to reject the claims under 102 and 103. EX1004, 803-810. Patentability of claim 1 over those same references was further discussed in an examiner interview. EX1004, 217. Then, in response to the rejection, Patent Owner amended claim 1, explaining that claim 1 is patentable over Flusberg and the ’338 patent. EX1004, 206-209. Following Patent Owner’s amendment and arguments, the examiner allowed claim 1 without further comment. EX1004, 174-178. As such, Flusberg and the ’338 patent were considered during prosecution with respect to the patentability of claim 1 of the ’794 patent.

Second, *Becton, Dickinson* factor b—the cumulative nature of the asserted art and the prior art evaluated during examination—favors denial. The third reference in Ground 1, the PacBio-BAM, is cumulative to the disclosure of Flusberg and the '338 patent and adds nothing to the patentability determination here. As Petitioner's own expert explains, "a BAM file is an industry standard file format" (EX2002, ¶52), and the PacBio-BAM is simply "the standard data structure in which PacBio SMRT sequencers store data" (EX1002, ¶81) that is "generated in the normal course of SMRT sequencing" (EX1002, ¶52).

During prosecution, the examiner recognized that both Flusberg and the '338 patent disclose using data generated in the normal course of SMRT sequencing, in other words, information cumulative to the PacBio-BAM. Specifically, the examiner recognized that both Flusberg and the '338 patent disclose "performing single molecule, real time (SMRT) sequencing of a sample of a nucleic acid molecule and obtaining data for" the "identify of nucleotide," the "position of the nucleotide, the "width of the pulse corresponding to each nucleotide" and the "interpulse duration (IPD)." as follows:

Flusberg et al teach detecting modifications of a nucleotide in a nucleic acid molecule comprising:

(a) performing single-molecule, real-time (SMRT) sequencing of a sample of nucleic acid molecule and displaying in a window (i.e. a computer screen or piece of paper) the data for:

- identity of the nucleotide T, C, G, and A in a sequence comprising at least 8 consecutive nucleotides (Figure 1);
- the position of the nucleotide;
- the width of pulse corresponding to each nucleotide;
- the interpulse duration (IPD) representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

EX1004, 803, 806.

The '338 patent further explains that the data generated in the normal course of SMRT sequencing is provided in a computer readable data structure like the PacBio-BAM. For example, the '338 patent recites that “[t]he systems of the invention also typically include information processors or computers operably coupled to the detection portions of the systems, in order to store the signal data obtained from the detector(s) on a computer readable medium....” EX1007, 71:49-53. The '338 patent also recites that the “computers also typically include software that analyzes the raw signal data, identifies signal pulses that are likely associated

with incorporation events, and identifies bases incorporated during the sequencing reaction, in order to convert or transform the raw signal data into user interpretable sequence data (see, e.g., Published U.S. Patent Application No. 2009-0024331, the full disclosure of which is incorporated herein by reference in its entirety for all purposes).” *Id.* 71:62-72:3. As such, the PacBio-BAM adds nothing to the disclosures of, and is cumulative to, Flusberg and the ’338 patent. EX2002, ¶131.

The prosecution of the ’794 patent further confirms that the PacBio-BAM is cumulative to Flusberg and the ’338 patent. Initially, the examiner rejected original claim 1 over both Flusberg and the ’338 patent, asserting that the claimed “received” data, which in original claim 1 included the “input data structure,” was disclosed in both references. EX1004, 803-810. Original claim 1 recited, in part:

receiving an input data structure, the input data structure corresponding to a window of nucleotides sequenced in a sample nucleic acid molecule, wherein the sample nucleic acid molecule is sequenced by measuring pulses in an optical signal corresponding to the nucleotides, the input data structure comprising values for the following properties:

for each nucleotide within the window:

an identity of the nucleotide,

a position of the nucleotide with respect to a target

position within the respective window,

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

EX1004, 1090.

In response, Patent Owner amended original claim 1 to clarify that the claimed “input data structure” is not simply the data generated in the normal course of SMRT sequencing, *e.g.*, the PacBio-BAM. Instead, the “input data structure” is created from “receiving data acquired by,” for example, PacBio’s SMRT sequencing. Specifically, claim 1 was amended to delete “receiving an input data structure,” and instead to recite “(a) receiving data acquired by...” and then “(b) creating an input data structure....” Amended claim 1 recited, in part:

1. (Currently Amended) A method for detecting a modification of a nucleotide in a nucleic acid molecule, the method comprising:

(a) receiving an input data structure, the input data structure corresponding to a window of nucleotides sequenced in a sample nucleic acid molecule, wherein the sample nucleic acid molecule is sequenced by receiving data acquired by measuring pulses in an optical signal corresponding to [[the]] nucleotides sequenced in a sample nucleic acid molecule, the input data structure comprising and obtaining, from the data, values for the following properties:

for each nucleotide ~~within the window~~:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid molecule ~~with respect to a target position within the respective window,~~

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;

(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;

(c) inputting the input data structure into a model, the model trained by:

* * *

EX1004, 191-92.

In summary, the amendment confirms that simply having the data generated by SMRT sequencing, *e.g.*, the PacBio-BAM, does not disclose or render obvious creating an “input data structure” having the claimed combination of features. Of course, the sequencing data was available in a computer readable format in both

Flusberg and the '338 patent. But, as the examiner recognized, the relevant question is whether Flusberg and the '338 patent disclose or render obvious receiving that SMRT sequencing data and then creating an input data structure having claimed attributes, namely:

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;

EX1001, claim 1.

The examiner determined that Flusberg and the '338 patent do not disclose or render that combination obvious. Specifically, following an interview discussing the proposed claim amendments and Flusberg and the '338 patent, the examiner stated that while “Flusberg displays various widths of pulses in their sequence window,” Flusberg does not appear to “measure the widths as a data point to be combined with interpulse duration, nucleotide identity and position in determining the presence of a mutation.” EX1004, 217. The Examiner referred to that same reasoning for nonobviousness with respect to the '338 patent. *Id.* The full statement is as follows:

35 U.S.C. 102

Examiner stated the method of Flusberg displays various widths of pulses in their sequence data in a window, however it does not appear that they measure the widths as a data point to be combined with interpulse duration, nucleotide identity and position in determining the presence of mutations. Examiner will consider and compare future claim amendments with the prior art.

35 U.S.C. 103

See comments in 102 above.

EX1004, 217.

Next, *Becton, Dickinson* factor (d)—the extent of the overlap between the arguments made during examination and the manner in which petitioner relies on the prior art or patent owner distinguishes the prior art—favors denial. As explained above, here Petitioner makes the same arguments considered and rejected during prosecution, namely whether Flusberg and the '338 patent render obvious “creating an input data structure” having the claimed combination of features. During prosecution, the examiner found that Flusberg and the '338 patent did not disclose those limitations and allowed the claims over those references.

Petitioner attempts to manufacture a new issue for the PacBio-BAM by misrepresenting Patent Owner's infringement contentions in a related litigation in which Patent Owner is suing Petitioner for infringing the '794 patent. Specifically, Petitioner claims that the infringement contentions purportedly assert that the PacBio-BAM is the claimed “input data structure.” Petition, 20.

But as explained above, the infringement contentions contain no such allegation. Instead, the contentions allege that the claimed input data structure is

created from the data in a PacBio-BAM file, not that it is a PacBio-BAM file.

EX1014, 25 (“The HiFi reads in the BAM file are then used to create the data structure of the claimed element [1.b], i.e., an input data structure, for claim element [1.c].”)

Thus, Petitioner’s assertion that the PacBio-BAM is new based on the infringement contentions is premised on a misrepresentation of the infringement contentions and should be rejected.

B. The Petition has not demonstrated that the Office erred in a manner material to the patentability of the challenged claims.

First, *Becton, Dickinson* factor (c)—the extent to which the asserted art was evaluated during examination, including whether the prior art was the basis for rejection—favors denial. As set forth above, Flusberg and the ’338 patent were both extensively considered during prosecution. The examiner rejected original claim 1 over both references (EX1004, 803-810), and Patent Owner amended claim 1 and explained why claim 1 is patentable over both references, both in writing (*id.*, 191-210) and during an examiner interview (*id.*, 217).

Second, *Becton, Dickinson* factor (e)—whether petitioner has pointed out sufficiently how the examiner erred in its evaluation of the asserted prior art—favors denial. Petitioner’s first argument that allowance of the ’794 patent was material error is premised on Petitioner’s misrepresentation of the Patent Owner’s infringement contentions. As explained above, the infringement contentions do

not admit or assert that the PacBio-BAM is the claimed “input data structure.”

Petitioner’s next argument that the examiner committed an error by confusing rejections under 102 and 103 for Flusberg and the ’338 patent misunderstands the examiner’s interview summary. In summarizing an interview in which the examiner and Patent Owner discussed Flusberg and the ’338 patent and possible claim amendments, under the heading “35 U.S.C. 102” the examiner stated that while “Flusberg displays various widths of pulses in their sequence data in a window,” Flusberg does not appear to “measure the widths as a data point to be combined with interpulse duration, nucleotide identity and position in determining the presence of mutations.” EX1004, 217. And regarding the 103 rejections, under the heading “35 U.S.C. 103” the examiner further stated “see comments in 102 above.” *Id.* The examiner’s statements do not suggest, let alone demonstrate, that the examiner confused application of 102 or 103. To the contrary, the summary confirms that the examiner recognized the relationship between Flusberg and the ’338 patent and that both include the same deficiency, namely, neither disclose or render obvious the claimed combination of features for the “input data structure.” Regardless, as the Board found in *LG Electronics v. Image Processing Tech.*, though “[s]tandards for anticipation and obviousness are distinct, [a]dvancing a new statutory basis...does not demonstrate...error in prior consideration [of a reference].” 2023 WL 5030795, *9 (PTAB 2023) (denying

institution based on § 325(d)).

The examiner's interview summary further confirms that the Petition presents the same issue considered and overcome during prosecution. The examiner recognized that the standard output of PacBio SMRT sequencing, *e.g.*, the PacBio-BAM, includes various data in a structured format. But, the question for patentability is not was the information available in some structure. The question is does Flusberg and the '338 patent render obvious creating from the PacBio-BAM an "input data structure" having the claimed combination of features. In allowing claim 1, the examiner determined that neither reference discloses or renders obvious the claimed invention.

Petitioner's final argument regarding a statement by the examiner in a different patent application, a continuation of the '794 patent, is not relevant to evaluating if there was an error material in the '794 patent's prosecution. Regardless, the statement for the different patent application cited by Petitioner is not an error. Petitioner selectively quotes the first portion of the examiner's statement, replacing a comma with a period and omitting the remainder of the sentence. Petition, 25. The examiner's full statement characterizing the closest prior art of record—three U.S. patents including the '338 patent at issue here—is as follows:

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.

First, just the like the examiner found in allowing the '794 patent claims, that statement confirms that the '338 patent does not disclose creating an “input data structure” having the claimed combination of features. And second, that the '338 patent includes disclosure related to “pulse width,” does not mean that the '338 patent “teaches measuring a pulse width value of a specifically identified nucleotide at an identified position to determine the methylation status.”

Petitioner's citation to excerpts from the '338 patent and Flusberg related to pulse width does not demonstrate an error material to prosecution.

Finally, *Becton, Dickinson* factor (f)—the extent to which additional evidence and facts presented in the petition warrant reconsideration of the prior art or arguments—favors denial. Petitioner makes no arguments regarding this factor. Petitioner does reference its expert's declaration and other references as new under

the first prong above. But, just like the petition, the report uses hindsight to second guess the precise issues addressed during prosecution. Regardless, this factor favors denial or is neutral. *See Volkswagen Group of America, Inc. v. StatosAudio, Inc.*, 2021 WL 6340178 at *6 (PTAB 2021) (finding factor (f) “at most, neutral” based on newly presented expert testimony and denying institution). Regarding the additional references cited in Petitioner’s expert declaration and Petition, as the Board found in *Sandoz Inc. v. Acerta Pharma B.V.*, “Petitioner’s reliance in the Petition on references that are not part of the prosecution record considered by the Examiner, and are not even asserted against the claims in Petitioner’s asserted grounds...fails to demonstrate ‘that the Office erred in a manner material to the patentability of the challenged claims.’” 2023 WL 5167505 at *8 (PTAB 2023) (internal quotes omitted).

For the foregoing reasons, Patent Owner requests that the Board exercise its discretion and deny the Petition.

IV. CLAIM 1 IS NOT OBVIOUS OVER GROUND 1.

The ’794 patent has one independent claim, claim 1. Claims 2-19 depend, either directly or indirectly from claim 1. Because all the claims require the input data structure of claim 1, all grounds of the Petition are premised on the combination of the PacBio-BAM, Flusberg and the ’338 patent. And because the combination of PacBio BAM, Flusberg and the ’338 patent do not teach or suggest

the input data structure of claim 1, Grounds 1-4 fail.

The framework for determining obviousness requires consideration of the scope and content of the prior art, differences between the prior art and the patent claims, the level of ordinary skill in the art, and any secondary considerations.

KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398, 406 (2007), citing *Graham v. John Deere Co.*, 383 U.S. 1, 15-17 (1966).

Obviousness must be considered based on the claim as a whole. *Ruiz v. A.B. Chance Co.*, 357 F.3d 1270, 1275 (Fed. Cir. 2004) (explaining that “[t]he ‘as a whole’ instruction in title 35 prevents evaluation of the invention part by part.... This form of hindsight reasoning, using the invention as a roadmap to find its prior art components, would discount the value of combining various existing features or principles in a new way to achieve a new result—often the very definition of invention.”).

So, even “in the case of a combination of references that together disclose all the limitations of the claimed invention,” there must be an “apparent reason to combine the known elements in the fashion claimed by the patent at issue” and a “reasonable expectation of success” to a POSITA in pursuing that combination. *Los Angeles Biomedical Rsch. Inst. at Harbor-UCLA Med. Ctr. v. Eli Lilly and Co.*, 849 F.3d 1049, 1064 (Fed. Cir. 2017) (cleaned up) (citing *KSR*, 550 U.S. at 418).

Moreover, the inventor's own path itself never leads to a conclusion of obviousness; that is hindsight and expressly made irrelevant to patentability by statute. *Millennium Pharmaceuticals, Inc. v. Sandoz Inc.*, 862 F.3d 1356, 1367 (Fed. Cir. 2017) (citing 35 U.S.C. § 103(a) (“Patentability shall not be negated by the manner in which the invention was made.”)).”

Prior art teaches away when “a person of ordinary skill, upon reading the reference, would be discouraged from following the path set out in the reference, or would be led in a direction divergent from the path that was taken by the applicant.” *Arctic Cat Inc. v. Bombardier Recreational Products Inc.*, 876 F.3d 1350, 1360 (Fed. Cir. 2017), citing *Allergan, Inc. v. Sandoz Inc.*, 796 F.3d 1293, 1305 (Fed. Cir. 2015).

Here, the Board should not second guess the Examiner’s determination that Flusberg and the ’338 patent do not disclose “measur[ing] the widths as a data point to be combined with interpulse duration, nucleotide identity and position in determining the presence of a mutation” and thus do not render obvious claim 1. EX1004, 217.

First, the assertion that the PacBio-BAM is an “off the shelf” “input data structure” as claimed is factually incorrect. EX2002, ¶126. Patent Owner’s infringement contentions do not allege that the PacBio-BAM is the “input data structure.” To the contrary, a POSITA would understand that the PacBio-BAM

does not include “a target position” and cannot be used as “input data structure” to train a model because it includes superfluous data and is not a consistent length necessary to train a model. EX2002, ¶¶127, 130. That the PacBio-BAM references detecting base modifications, does not suggest to a POSITA how to create an input data structure or to include the combination of features of the claimed input data structure. EX2002, ¶129.

Second, Flusberg, the ’338 patent, and the PacBio-BAM do not render obvious the claimed “input data structure” including a window of nucleotides” and “for each nucleotide within the window” the claimed combination of “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.” Both Flusberg and the ’338 patent do not disclose and instead teach away from including “identify of the nucleotide” and both references teach away from including “the width of the pulse corresponding to the nucleotide” as explained by several of the same authors of the two references in later publications. EX2002, ¶¶67, 75, 78. That the underlying data was available in a PacBio-BAM file, does not render the claimed combination of data obvious.

Petitioner’s argument “the ’794 patent simply follows teaching[s] in the prior art” citing to the ’794 patent itself (Petition, 14-17) cannot provide the required motivation to arrive at the claimed invention. It does not matter whether

the inventors reached their invention after an exhaustive study of the prior art, or developed the invention in complete isolation. *See Life Techs., Inc. v. Clontech Labs., Inc.*, 224 F.3d 1320, 1325 (Fed. Cir. 2000).

Accordingly, because Petitioner has failed to demonstrate that claim 1 is obvious under Ground 1, Patent Owner requests that the Petition be denied.

A. PacBio-BAM does not disclose or render obvious the “input data structure” or its combination of features recited in claim 1.

Contrary to Petitioner’s assertion, Patent Owner’s infringement contentions do not assert that the PacBio-BAM is the claimed “input data structure.” As explained above, the infringement contentions recite that the “input data structure” having the claimed combination of features is “created” from the PacBio-BAM. It is not the PacBio-BAM. Ground 1 fails for that reason alone.

Regardless, contrary to Petitioner’s Expert’s assertion that the PacBio-BAM is an “off the shelf option” for the input data structure (EX1002 ¶ 82), a POSITA would understand that while the PacBio-BAM includes information from which the input data structure could be created, the PacBio-BAM cannot be the claimed “input data structure” and does not suggest or provide motivation to select various data included in the PacBio-BAM to create the claimed input data structure.

EX2002, ¶¶126-127, 130-131.

1. The PacBio-BAM contains extraneous information that would render training a model inoperable.

First, the PacBio BAM cannot be the “input data structure” of claim 1 because it is in a compressed binary format that cannot be read by a machine learning model and comprises text data, not an input data array necessary for training a model. EX2002, ¶126. ; EX1011, 1 (“a binary, compressed, record-oriented container format for raw or aligned sequence reads.”). The PacBio-BAM file is simply the digital form of the SMRT sequencing data which is “generated in the normal course of SMRT sequencing.” EX1002, ¶52. Because the PacBio-BAM storage format is incompatible as input data structure for training a model, it is not a “preexisting off-the-shelf option” of an input data structure for a trainable model. EX2002, ¶¶126-127.

Moreover, the PacBio BAM contains not only base calling data and kinetics data obtained from SMRT sequencing, but also other numeric and non-numeric data that must be parsed out to create an “input data structure” to train a model, including for example, SAM Flags, subread name, context flag, number of passes, base quality, position end, position start, read quality, signal-to-noise, zmw number and read group. EX2002, ¶126. EX1011, 3-7. Indeed, some of that data is labeled “Mandatory Items” in the PacBio-BAM as shown below:

Mandatory items:

Key	Value spec	Value example
READTYPE	One of ZMW, HQREGION, SUBREAD, CCS, SCRAP, or UNKNOWN	SUBREAD
BINDINGKIT	Binding kit part number	100236500
SEQUENCINGKIT	Sequencing kit part number	001558034
BASECALLERVERSION	Basecaller version number	2.1
FRAMERATEHZ	Frame rate in Hz	100
CONTROL	TRUE if reads are classified as spike-in controls, otherwise CONTROL key is absent	TRUE

EX1011, 4.

2. The PacBio-BAM does not include a “target position.”

Second, the PacBio-BAM includes no “target position within a window” let alone “a position of the nucleotide with respect to a target position within the window” as required by claim 1.

Claim 1 recites that the “input data structure” requires a “position of the nucleotide with respect to a target position within the window.” Claim 1 further recites “(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.” As such, the “target position within the window” of claim 1 is the nucleotide being interrogated for modification in step (d). Because the PacBio-BAM is the standard output of SMRT sequencing that can be used for many purposes, it does not include or identify a “target position within a window” or a “a position of the nucleotide with respect to a target position within the window” as required by

claim 1. It is simply the sequence data. EX2002, ¶130

Petitioner plucks the word “target” from the PacBio-BAM to allege that the “target position” is disclosed. Petition, 41-42. But contrary to Petitioner’s assertion, a POSITA would understand that the PacBio-BAM’s recitation of the word “target” in sequence alignment is unrelated to and does not disclose the claimed “target position within the window.” EX2002, ¶130. Instead the term “target” in the PacBio-BAM refers to the reference sequence, *i.e.*, “target sequence” to which the SMRT sequence read is aligned to generate an aligned BAM file. It is not a target position for base modification detection, and does not suggest creating an “input data structure” having “a position of the nucleotide with respect to a target position within the window” as claimed. EX2002, ¶130.

3. The PacBio-BAM cannot be used to train a model.

The PacBio-BAM cannot be the claimed “input data structure” because as set forth above, claim 1 further requires in step (c), a “model trained by” “receiving a first plurality of first data structures” where “each first data structure comprising values for the same properties as the input data structure.” As such, a POSITA would understand that the “input data structure” must have the same structure as the “first plurality of first data structures” used to train the model. EX2002, ¶38; EX1002, ¶130. But, sequence data in the format of the PacBio-BAM is of varying lengths, and thus cannot be used to train a model as required by claim 1. EX2002,

¶128.

Specifically, a PacBio chip for SMRT sequencing has tens of thousands to millions of wells in which tens of thousands to millions of nucleic acid molecules are sequenced. In 2011, the PacBio chip had 45,000-75,000 wells. EX1026, 128. A recent PacBio high density chip has 25 million wells. EX2027, ¶128. Whether the chip has millions or tens of thousands of wells, DNA sequences of various sizes are sequenced and stored in the PacBio-BAM file. The size of each sequence in each well is random; most will be of different lengths. EX2002, ¶128. Because the PacBio-BAM file includes all the sequence data from the PacBio chip, the PacBio-BAM file will have tens of thousands to millions of sequences of varying lengths. Even assuming that the extraneous data was first parsed and “a position of the nucleotide with respect to a target position within the window” was added to the PacBio-BAM, the stored sequences cannot be used to train a model because they have varying lengths. EX2002, ¶128.

Petitioner’s citation to the PacBio-BAM’s recitation that its file format “[p]rovides a substantial disk-space savings without affecting important production use cases (base modification detection)” does not mean that the PacBio-BAM is the claimed “input data structure” or suggest to a POSITA what combination features in the PacBio-BAM should be used to create an input data structure. EX2002, ¶129. Instead, that sentence indicates that data is preserved in the BAM

from which modification can be detected. EX2002, ¶129. In other words, just like in Flusberg, the standard output of SMRT sequencing is preserved, such that a POSITA must then figure out what to include in an “input data structure” to detect base modification. Because the PacBio-BAM is the standard output of SMRT sequencing, a POSITA must look elsewhere for motivation to select a particular combination of features to include in an “input data structure” for base modification detection.

B. Flusberg and the '338 patent do not disclose or render obvious the “input data structure” or its combination of features recited in claim 1.

1. Flusberg does not disclose an “input data structure” including the “identity of a nucleotide” and teaches away from including pulse width.

The claimed “input data structure” requires “for each nucleotide within the window: the nucleotide identity, nucleotide position with respect to a target position within the window, the nucleotide PW, and the IPD.” Flusberg does not disclose or render obvious that claimed “input data structure” at least because Flusberg does not disclose, and instead teaches away from including, “identity of the nucleotide.” EX2002, ¶78. Flusberg also teaches away from using pulse width. EX2002, ¶¶67, 75.

First, Petitioner’s own expert admits that Flusberg avoids using the identity of the nucleotide as a feature for an input data structure as follows: “Flusberg’s

specific example avoids using sequence context as an express input variable because it is held constant – Flusberg’s principal component analysis model is trained on a particular sequence that is varied in methylation state, but not in nucleotide sequence.” EX1002, ¶107. In other words, a POSITA would understand that Flusberg not only does not use nucleotide identity, but teaches away from using nucleotide identity in an “input data structure” because Flusberg’s method expressly avoids its use as an input data. EX2002, ¶78. Flusberg does not use nucleotide identity as a feature for an input data structure because nucleotide identity is a constant in the Flusberg experiments, not a variable. EX2002, ¶78.

Because identity of the nucleotide was a constant in Flusberg, Flusberg uses the term “sequence context” to refer to a window of nucleotides around a known methylated nucleotide, or as a hindsight observation that methylation affected polymerase kinetics of other nucleotides around the base. EX2002, ¶¶63, 66. For example, when describing the design of synthetic DNA templates for kinetic data analysis having identical primary sequence with different methylation sites, mA, mC, and hmC, Flusberg discloses “[i]n all cases, mA was located in a GATC sequence context, and mC and hmC were located in C-G context.” EX1006, 462. The term “sequence context” does not refer to an input data. EX2002, ¶63.

Another example of Flusberg’s use of “context” refers to GATC “context,” which showed a higher IPD ratio associated with “expected” adenine methylation

in GATC “context.” EX1006, Figure 5, Table 1, Supplementary Figure 5; EX2002,

¶77. Again, that “context” does not refer to an input data. EX2002, ¶77.

Table 1 | IPD ratios for each fosmid GATC motif

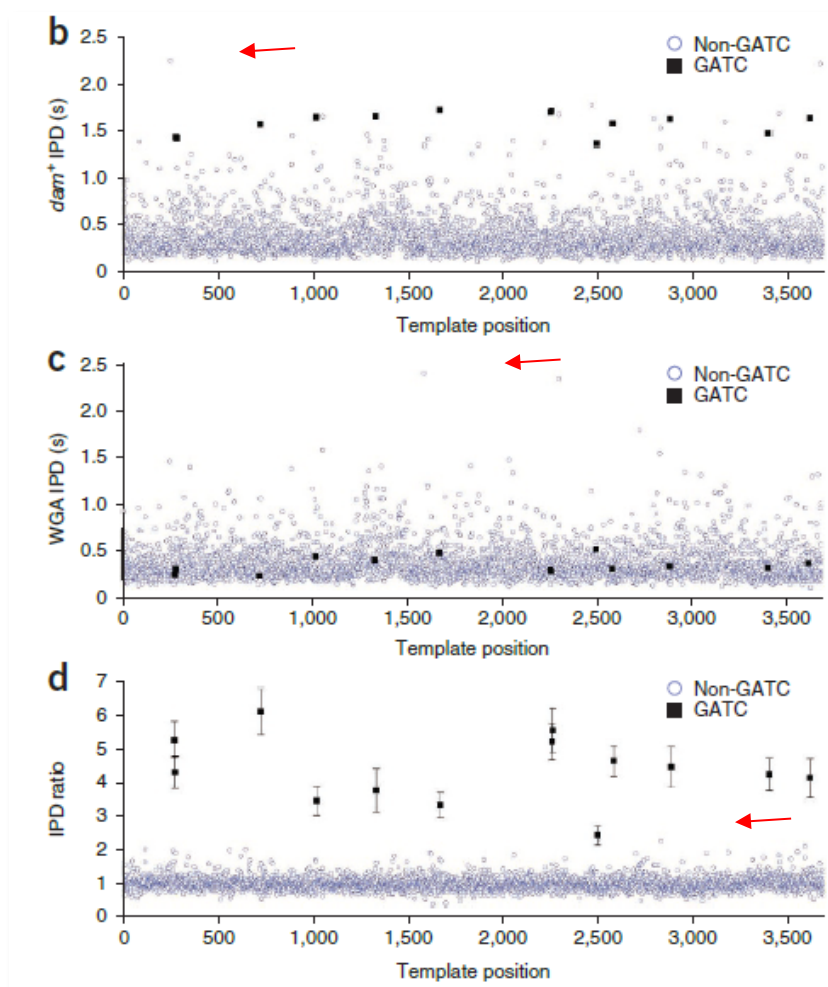
Position	Sequence	IPD ratio	P value
273	TGCCATGATCTAGATC	5.28	2.84×10^{-18}
279	GATCTAGATCATCGTG	4.32	4.41×10^{-16}
720	TTCTATGATCAGGGAG	6.12	7.00×10^{-21}
1015	GCGTGGGATCTGTATG	3.46	6.25×10^{-13}
1329	TATCAGATCTCATT	3.78	2.35×10^{-12}
1668	TAGTTGGATCAAGAGA	3.33	2.61×10^{-13}
2256	CTTTTGGATCAGATCC	5.22	1.70×10^{-19}
2261	GGATCAGATCCAATTA	5.56	1.43×10^{-22}
2499	CAGATGGATCAATCAA	2.43	4.44×10^{-7}
2583	ATTTTGGATCTAGTTT	4.65	2.00×10^{-21}
2887	ATTCGGATCTCCACA	4.46	1.57×10^{-14}
3402	CCTCAAGATCATCATC	4.24	2.04×10^{-15}
3619	GCCAGCGATCATATT	4.13	2.05×10^{-10}

For each GATC motif in the 3.7 kb fosmid subsection (Fig. 5), the local sequence context IPD

EX1006, 464.

That experiment used a *C. elegans* fosmid subsection isolated from a DNA adenine methyltransferase-positive (*dam*⁺) *E. coli* strain was used for the methylated sample. The same fosmid subsection was subjected to whole-genome amplification (WGA) to create unmethylated control samples. EX1006, 464. The samples were sequenced and the IPD values plotted GATC motifs (solid square); non-GATC motifs (open circle). EX1006, Figure 5(b)&(c). GATC motifs in the methylated samples had significantly higher IPD values (~1.5 s) than GATC motifs in the control samples (<~0.5 s). EX1006, Figures 5(b)&(c). Non-GATC motifs in both samples registered IPD values (red arrows) comparable to the methylated

GATC motifs. By dividing the IPD values of methylated samples (*dam*⁺, Figure 5(b)) over control (WGA, Figure 5(c)), the resulting IPD ratios improved the separation of methylated GATC motifs over non-GATC motifs. EX2002, ¶74.



EX1006, 464 (annotated).

As some of Flusberg's co-authors later explained, IPD ratios "[t]o eliminate the context-dependent fluctuation of the IPD values, we calculated the IPD ratio (IPDR) on each genomic position as previously described (Flusberg *et al.*, 2010)."

EX2005, 2912, citing EX1006. As such, Flusberg teaches away from including

identity of the nucleotide in the input data structure because IPD ratios eliminate any reason to do so. EX2002, ¶75. In other words, because Flusberg discloses using IPD ratios, a POSITA would understand that identity of the nucleotides is not needed. *Id.*

Second, other co-authors further explained that Flusberg teaches away from including in an “input data structure” including “the width of the pulse corresponding to the nucleotide [*i.e.*, PW]” as claimed, as follows:

Given the high degree of regulation involved in the base incorporation process, PW is not as sensitive to shifts in sequence context induced by modifications as IPD (Flusberg et al. 2010). Therefore, for the purposes of developing an initial model of kinetic variation, we restrict attention to IPDs.

EX1010, 130.

Accordingly, the coauthors of Flusberg confirm that Flusberg does not disclose or motivate a POSITA to include “identity of the nucleotide” and pulse width in an input data structure.

2. The '338 patent does not disclose an “input data structure” including the “identity of a nucleotide” and teaches away from including pulse width.

First, like Flusberg, the '338 patent uses sequence or local “context” to refer to the positions before and after a modified base, recognizing that polymerase kinetics are affected by the surrounding positions. EX2002, ¶¶82, 102, 105-116. As such, the '338 patent does not include “identity of the nucleotide” as a feature

for an “input data structure.” EX2002, ¶¶82, 102, 105-116.

Petitioner cites to a passage from the '338 patent as disclosing use of “identity of the nucleotide.” Petition, 29. But the passage does not include “identity of the nucleotide” and instead confirms that the term “context” means positions before and after a modification, not the identity of the nucleotide at those positions. EX2002, ¶105. The full citation reads as follows:

In certain aspects, the invention provides a general-purpose approach to discriminating between Me⁺ and Me⁻ using features in a real-time sequencing-by-synthesis trace comprising signals emitted during the incorporation of optically detectable nucleotides into a nascent strand by a polymerase enzyme.... A first stage of this approach includes the development of a classifier for distinguishing methylated from unmethylated cytosines in a nucleic acid template. A set of features is measured for every pulse (discrete event or signal) in the trace. For example, a set of measurable features might be {pulse width, pulse duration, pulse height, pulse amplitude variability}. Call the values of these features $f_1^i, f_2^i, \dots, f_k^i$, for pulse i . For each cytosine pulse in the methylated and unmethylated template data sets (which may or may not be restricted to CpG), tabulation is performed for the local pulse features

$$\vec{f} = \{f_1^{i-3}, f_2^{i-3}, \dots, f_k^{i-1}, f_1^i, f_2^i, \dots, f_k^i, \dots, f_1^{i+3}, f_2^{i+3}, \dots, f_k^{i+3}\}.$$

In this example, a local context extending 3 pulses to the left and right of the pulse of interest is assumed, but this context size is flexible and,

in certain embodiments, can be application-specific.
EX1007, 66:46-67:4.

While the above passage recites a “set of features,” those features do not include, let alone suggest including, “identity of the nucleotide” as claimed. EX2002, ¶105. Instead, the passage refers to a “local context” that is “extending 3 pulses to the left and right of the pulse of interest is assumed, but this context size is flexible....” EX1007, 67:1-3. That passage makes clear that “local context” means relative position, not identity of the nucleotide at that position.” EX2002, ¶105.

Moreover, the “set of features” includes a laundry list of possible features reciting that “[f]or example, a set of measurable features might be {pulse width, pulse duration, pulse height, pulse amplitude variability}.” EX2002, ¶105. Again, while the ’338 patent leaves to POSITA to figure out what features to include, “identity of the nucleotide” is not one of those features. EX2002, ¶105.

Neither the Petitioner, nor its expert, point to any disclosure in the ’338 patent that nucleotide identity is used for detecting base modification, let alone as a feature to be included in an input for modification determination. EX2002, ¶¶105-116. To the contrary, as the co-inventors explained six months after the issuance of the ’338 patent, IPD ratios “eliminate the context-dependent fluctuation of the IPD values.” *Id.*, EX2016, 2912 (citing EX1006).

Regarding the other features, the '338 patent does not teach a POSITA to use the combination of IPD and PW together with the other claimed features in an “input data structure.” EX2002, ¶¶105, 117-124. A POSITA would understand that the selection of features to combine in a data structure is a fundamental problem in training a model. EX2002, ¶26. Because some data carry redundant information and other data does not have any predictive information, these features can introduce excess noise, the selection of the appropriate features as data inputs for a trainable model is unpredictable. EX2002, ¶26. To select appropriate features, the predictive power of the features needs to be assessed in combination with the training data, model choice, and the task. EX2002, ¶26. None of that was shown in the '338 patent. EX2002, ¶105.

Like Flusberg, all the working examples in the '338 patent are simple idealized experiments using nucleic sequences with contrived patterns of modified and unmodified bases along with control sequences having the same primary sequence but without any modification. EX2002, ¶102. Every experiment compares the IPD ratios between a known modified nucleic acid sequence and an unmodified nucleic acid sequence with identical nucleotide sequence. EX2002, ¶¶82-87. EX1007, 75:32-76:64 (5-methylcytosine), EX2002, ¶¶88-89, 76:65-77:62 (N⁶-methyadenosine), EX2002, ¶¶90-92, 77:63-78:17 (5-hydroxymethylcytosine), EX2002, ¶¶93-95, 78:18-43 (8-oxoguanosine), EX2002,

¶¶96-97, 78:42-79:19 (Ribonucleosides within DNA), EX2002, ¶¶98-99, 79:19-46 (Glucose-Modified 5-hydroxymethylcytosine), EX2002, ¶¶100-101, 79:47-80:35 (Modified Bases in RNA Templates). While a few new base modifications are tested, the data is largely duplicative of Flusberg showing that IPD is strongly associated with base modifications. EX2002, ¶79. The '338 patent does not include a single experiment with *de novo* detection of DNA modification, *i.e.*, an unknown modified base. EX2002, ¶102, 134.

And, like Flusberg there is scant data in the '338 patent on the effect of PW on base modifications. Only two of the eight experiments measure PW ratios. EX2002, ¶91, 94. The first is the 5-hydroxymethylcytosine experiment, which appears to be the same Flusberg data, just plotted at a different scale², (see Y axis). EX1006, 462, EX1007, 77:63-78:17; EX2002, ¶92.

² Notably, the 5-methylcytosine PW data and the principal component analysis experiments including Fig.3 of Flusberg are omitted from the '338 patent, which was filed a continuation-in-part application November 12, 2010, about six months after the Flusberg on-line publication of May 9, 2010. EX1006, 461, EX1007, face page.

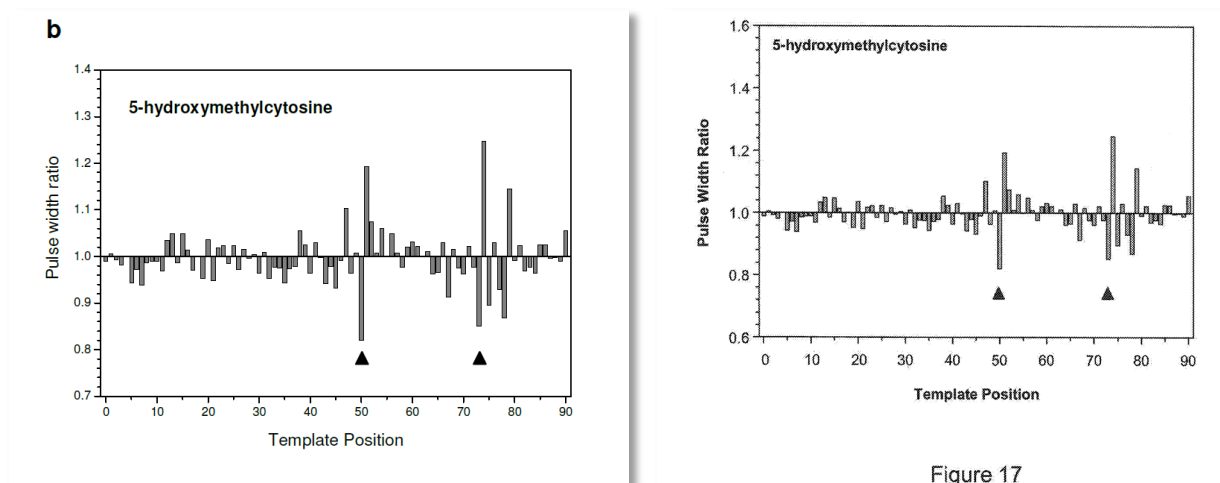


Figure 17

The other experiment to measure PW ratios was 8-oxoguanosine. Like 5-methylcytosine, IPD ratios were strongly correlated with base modification. (“8-oxoG altered IPD significantly” “[s]ome positions saw an increase in IPD by a factor of as much as 6.5X.”). EX1007, 78:30-33, EX2002, ¶94. Yet, for PW the correlation was not strong, the PW ratios at and around the modified base had a low dynamic range compared to the IPD which had a substantially higher signal-to-noise ratio. EX2002, ¶94. As the ’338 patent’s coinventors later explained, a person of skill in the art reviewing the ’338 patent data would conclude “PW is not as sensitive to shifts in sequence context induced by modifications as IPD.” Schadt, EX1010, 130. And, based on those teachings, just like Schadt, the motivation would be to “restrict attention to IPDs.” *Id.*

Lastly, Petitioner points to a now issued continuation of the ’338 patent as recognition by the Office that the ’338 patent discloses the “input data structure having the combination of features in claim 1. Petition, 28. But, the prosecution

of a continuation of the '338 patent including different claims and terms is not relevant to determining if Ground 1 renders claim 1 of the '794 patent obvious here. For the reasons set forth above it does not.

C. Petitioner provides no motivation to combine Flusberg in the three-reference combination of Ground 1.

Ground 1 is also fatally deficient because Petitioner fails to provide motivation to combine Flusberg with the other two references of Ground 1 to arrive at the claimed invention. Petitioner first alleges that Flusberg, in its entirety, is purportedly “incorporated by reference” in the '338 patent. Petition, 11. Then Petitioner treats Flusberg and the '338 patent as a single reference for the obvious combination of Ground 1, referencing the “Flusberg Publications” and arguing that the collective “Flusberg Publications” would be combined with the PacBio-BAM. Petition, 29. Petitioner makes no argument for why Flusberg and the '338 patent would be combined, let alone why a POSITA would selectively combine various aspects of each reference and include features not used in either reference to arrive at claim 1. The Petition should be denied for that reason alone.

“To incorporate material by reference, the host document must identify with detailed particularity what specific material it incorporates and clearly indicate where that material is found in the various documents.” *Zenon Environ. v. Filter Corp.*, 506 F.3d 1370, 1379 (Fed. Cir. 2008) (plain language “indicating that the disclosures are not incorporated in their entirety.”) In *Nautilus, Inc. v. Icon Health &*

Fitness Inc., the PTAB followed *Zenon* and rejected an assertion that “the use of the phrase ‘in its entirety’ indicates full rather than partial incorporation.” 2018 WL 6321395 at *18-*19 (PTAB 2018). In *Nautilus*, the application recited that “one example of...is set forth in U.S. Patent Application..., the disclosure of which is incorporated by reference herein in its entirety.” *Id.* The Board determined the phrase “the disclosure of which is incorporated by reference herein in its entirety” refers to the “one example,” not the entirety of referenced patent application.

Like in *Nautilus*, Flusberg is not incorporated in its entirety into the ’338 patent. Instead, the ’338 patent makes clear that only a few specific disclosures of Flusberg are incorporated in their entirety. In one passage, the ’338 patent incorporates Flusberg’s disclosure of the observed relationship of IPD and modifications as follows: “interpulse duration (IPD) was clearly lengthened by the presence of N⁶-MeA,” and “[f]urther details on the detection of N⁶-MeA, as well as other modifications..., are provided in Flusberg..., incorporated herein by reference in its entirety for all purposes.” EX1007, 77:26-34. In another passage, the ’338 patent incorporates Flusberg’s disclosure that “[c]losed circular DNA templates were constructed...using methods described elsewhere herein and in Flusberg..., incorporated herein by reference in its entirety for all purposes.” EX1007, 78:58-66. Both passages incorporate aspects of Flusberg, not the entirety of the reference. Indeed, incorporating the entirety of Flusberg in two separate

passages would be redundant and unnecessary, further confirming that the '338 patent intended to only incorporate those specific portions of Flusberg.

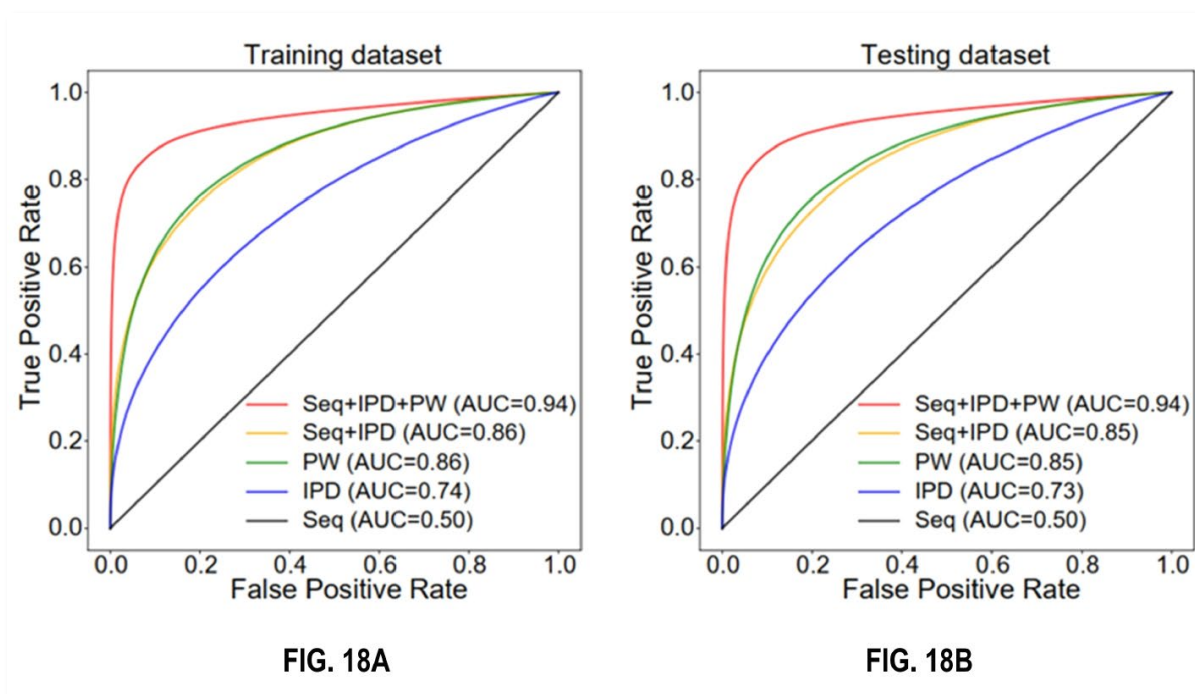
Regardless, even if the '338 patent incorporates the entirety of Flusberg—it does not—Petitioner still provides no motivation combine the various aspects of each reference to arrive at the claimed combination, which is insufficient as a matter of law. *See Los Angeles Biomedical Rsch. Inst.*, 849 F.3d at 1064.

D. Modification detection using the claimed input data structure provided unexpected results with higher accuracy compared to the cited prior art.

Nonobviousness may be established when an invention yielded more than predictable results. *Millennium Pharmaceuticals, Inc.*, 862 F.3d at 1368. “[W]hen unexpected results are used as evidence of nonobviousness, the results must be shown to be unexpected compared with the closest prior art.” *Id.* Unexpected results must have a “nexus” to the claims, i.e., a legally and factually sufficient connection between the evidence and the patented invention. *Amgen Inc. v. Sandoz Inc.*, 66 F.4th 952, 964 (Fed. Cir. 2023).

Flusberg and the '338 patent were cited by the examiner as the closest art of record. EX1004, 217. Based upon the disclosure of Flusberg and the '338 patent, a person of ordinary skill in the art would expect that IPD would have more predictive power than PW in detecting base modifications as confirmed by the coauthors in Schadt. EX1010, 130 (“restrict attention to IPDs.”); EX2002, ¶143.

An unexpected finding in the '794 patent is that the predictive power of PW is greater than IPD. EX2002, ¶143. And, the significantly improved predictive power from the combination of features that make up the input data structure of claim 1, identity of the nucleotide and relative position to the target (“Seq”), IPD, and PW is unexpected over the closest prior art. EX1001, 32:18-27. The '794 patent, FIG. 18 illustrates that result. EX2002, ¶143.



EX2001, Fig. 18.

E. Objective indicia strongly favors nonobviousness of claim 1.

“Objective evidence of nonobviousness includes: (1) commercial success, (2) copying, (3) industry praise, (4) skepticism, (5) long-felt but unsolved need, and (6) failure of others.” *Volvo Penta of the Americas, LLC v. Brunswick Corp.*, 81 F.4th 1202, 1215 (Fed. Cir 2023) (vacating and remanding because “the Board

failed to properly consider the evidence of objective indicia of nonobviousness.”). For objective indicia to be relevant, “there must be a nexus between the merits of the claimed invention and the objective evidence.” *Id.* at 1210 (citations omitted). That nexus can be satisfied by “by showing that the evidence of secondary considerations is the ‘direct result of the unique characteristics of the claimed invention.’” *Id.* (citations omitted).

Here, objective indicia strongly favor nonobviousness of claim 1. First, Petitioner’s own art relied upon in Ground 1 demonstrate a long felt and unsolved need for an improved and more accurate method of “detecting a modification of a nucleotide in a nucleic acid molecule,” as recited in claim 1.

Then, when Patent Owner’s invention in the ’794 patent was disclosed in an article published in a scientific journal, Tse et al., “Genome-wide Detection of Cytosine Methylation by Single Molecule Real-time Sequencing,” Proceedings of the National Academy of Sciences, 118(5) (2021) (“Tse”) (EX1030), Petitioner recognized and praised the Patent Owner’s invention. Specifically, Petitioner’s own president and CEO congratulated the inventors of the ’794 patent for their “development of a new method for highly accurate simultaneous determination of DNA sequence and CPG methylation in one go[.]” that will “help researchers explore the impact of epigenetic changes in humans and other organisms and has great potential in clinical research and eventually diagnostics in cancer and other

disease areas where methylation changes are known to be important disease markers.” EX2017, 5.

Petitioner then proceeded to copy the claimed invention. Petitioner’s infringement of the ’794 patent claim 1 is currently being litigated in a related case pending in the Norther District of California. EX2016.

Accordingly, Petitioner’s own actions attempting to solve the issue addressed in the ’794 patent, praising Patent Owner’s solution to Petitioner’s investors and the scientific community, and then copying that solution by infringing the ’794 patent are strong evidence of nonobviousness that should be accorded substantial weight. As such, Petitioner’s assertion in Ground 1 that claim 1 is obvious over Petitioner’s own prior art is outweighed by the objective indicia.

1. Petitioner’s long felt but unsolved need

The ’794 patent, which claims priority to August 2019, explains that “DNA methylation is essential for mammalian development and has notable roles” in many biological processes. EX1001, 1:52-58. The ’794 patent continued, “many human diseases have been found to be associated with aberrations of DNA methylation....” *Id.*, 1:61-63. As such, the ’794 patent recognizes that “there is a need to determine the modification of bases of nucleic acids, without” the drawbacks of prior art techniques, “e.g., bisulfite conversion.” *Id.*, 2:28-31.

That same need was previously recognized by Petitioner several years earlier in the two primary references of Ground 1 namely, Flusberg and the '338 patent. First, Flusberg, on its face was generated by PacBio, and the '338 patent is assigned to PacBio. *Compare* EX1006, 461 with EX1007, cover page. The two references include several common authors/inventors. *Id.* Flusberg notes that “[b]ecause of its key role in human health and disease, cytosine methylation is the most widely studied of the DNA modifications....” EX1006, 461. Flusberg further criticized “bisulfate treatment” and explained that “no high-throughput method has been demonstrated that allows the determination of primary sequence at the same time as methylation status.” *Id.* Like Flusberg the '338 patent further explains that “[r]esearch efforts aimed at increasing our understanding of the human methylome would benefit greatly from the development of a new methylation profiling technology that does not suffer from the limitations described above. Accordingly, there exists a need for improved techniques for detection of modifications in nucleic acid sequences, and particularly nucleic acid methylation.” EX1007, 3:7-13.

The failure of Flusberg and the '338 patent to meet that need was confirmed by several coauthors of both references a few years later. Both Flusberg and the '338 patent share several coauthors with Schadt (EX1010). Schadt recites that Flusberg discloses “the possibility of directly observing chemical modification or

damage events to nucleotides, and even discriminating between such events as a routine part of the SMRT DNA sequencing process.” EX1010, 130 (citations omitted). But, Schadt continues, “no statistical framework has been proposed to identify kinetic variation events from SMRT sequencing data that can serve as a surrogate for chemical modifications of DNA or RNA sequences.” EX1010, 130. In other words, according to some of the same authors of Flusberg and the ’338 patent, Flusberg discloses that it might be possible to detect modifications, but three years after Flusberg and the ’338 patent³, no actual method of doing so had been proposed. As set forth herein, Petitioner’s Ground 1 does not rely upon Schadt or any one of several other publications by Petitioner attempting to solve the issue because Flusberg and the ’338 patent directed a POSITA down a different path, focusing on IPD as opposed to PW or identity of the nucleotide as a feature for an input data structure. EX2002, ¶¶67, 75, 78. Indeed, PacBio researchers and collaborators followed that path. EX2002, ¶¶123, 124.

³ Note the parent application to the ’338 patent, which includes substantially the same disclosure as the ’338 patent, published in September 2010. EX2029, cover page.

2. Petitioner's praise for the claimed invention.

The nonobviousness of the '794 patent is further confirmed by Petitioner's praise for the invention of the '794 patent as described in Tse (EX 1030).

Specifically, representatives of Petitioner praised the invention of the '794 patent both when presenting to Petitioner's investors and to the scientific community.

In 2021, an article ("Tse") authored by several inventors of the '794 patent published that included an embodiment of claim 1 for direct determination of 5mC. *See* EX1030. Following publication, Patent Owner sent the article to Petitioner to negotiate a collaboration or licensing agreement related to the '794 patent's invention. EX2016, ¶15.

Petitioner then praised the invention. Specifically, Petitioner provided a link to the PNAS article on its own website. EX2016, ¶15; EX2019, 29. Then on February 10, 2021, Petitioner's president and chief executive officer, Christian Henry, praised the invention in a public earnings call stating that he would "like to share some customer highlights that really demonstrate the power and potential of PacBio HiFi sequencing technology." Mr. Henry continued, "we'd like to congratulate Dr. Dennis Lo and his colleagues at the Chinese University of Hong Kong[, the inventors of the '794 patent,] on the development of a new method for highly accurate simultaneous determination of DNA sequence and CPG methylation in one go, published in the PNAS Journal in January." EX2017, 5. In

Mr. Henry's own words, "this new method will help researchers explore the impact of epigenetic changes in humans and other organisms and has great potential in clinical research and eventually diagnostics in cancer and other disease areas where methylation changes are known to be important disease markers." EX2017, 5.

Petitioner's praise of the invention did not end with its president and CEO. In October 2021, at the 2021 annual meeting of the American Society of Human Genetics (2021 ASHG annual meeting), Petitioner's president of segment marketing, Jennifer Stone, highlighted "the great work coming out of Dennis Lo's lab at the Chinese University of Hong Kong earlier this year." EX2020 (video) and EX2023 (transcript). Ms. Stone continued, "[t]he team developed a methodology to directly examine 5-methyl-C at the whole genome level using the kinetic information from real-time HiFi sequencing without the need for bisulfite or enzymatic conversions. In the PNAS paper, Dr. Lo and his team described the development of algorithms that leverage HiFi sequencing information from both DNA strands to improve the detection of 5-methyl-C. The paper describes its application to human whole-genome sequencing and the benefits of characterizing the epigenome with long accurate reads. It allows epigenetic phasing and allelic differential methylation. So just like the phasing of genetic variants, the ability to detect the methylation status genome-wide with long HiFi reads also allows for long-range epigenetic phasing information." *Id.*

3. Petitioner copied the claimed invention as set forth in Tse.

Following Patent Owner sending Tse to Petitioner, Petitioner then copied the invention of the '794 patent. During same 2021 ASHG annual meeting in which Petitioner's president of segment marketing praised the invention of the '794 patent, PacBio's Associate Director of Product Marketing, Dr. Aaron Wenger, gave a presentation in which he stated that PacBio intended to make improvements to its sequencing platform that included the same benefits as the '794 patent. EX2024 (video) and EX2026 (transcript). Specifically, Dr. Wenger described the improvement as including "models that integrate kinetic information over multiple bases and passes of a molecule" that "can provide high precision and recall for 5mC from PacBio HiFi reads." *Id.*, 15:46–15:56. Dr. Wenger further explained that "there are tools available to call 5mC that are under development at PacBio and available to early access collaborators." *Id.*, 16:10–16:17.

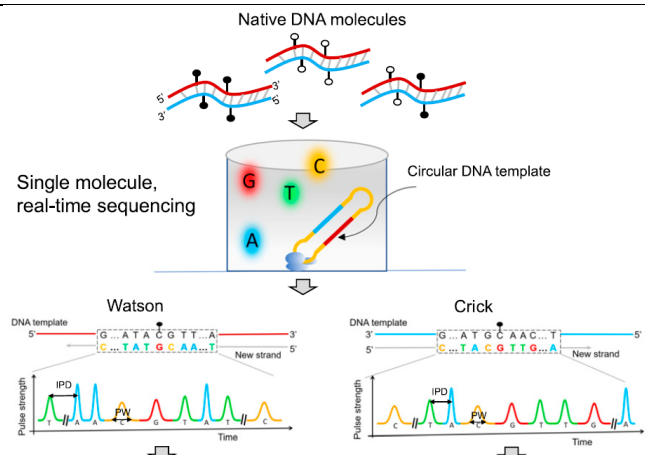
And sure enough, in April 2022 Petitioner released a new version of its SMRT Analysis Software suite—SMRT Link v11.0—including the new feature "New data utility to perform 5mC CpG Detection." EX2028, 2.

A detailed explanation of how Petitioner's SMRT Link v11.0 software infringes the '794 patent is found in EX2016, ¶¶ 34-63 and EX1014, 5-50.

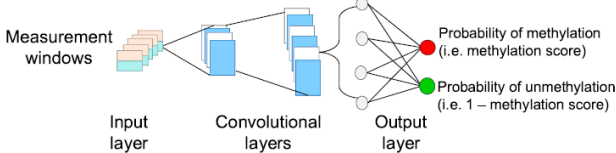
4. Nexus between Tse and claim 1.

Petition’s praise for and copying of Tse is the direct result of the unique characteristics of the claimed invention such that there is a nexus between the praise and copying and the objective indicia of nonobviousness.

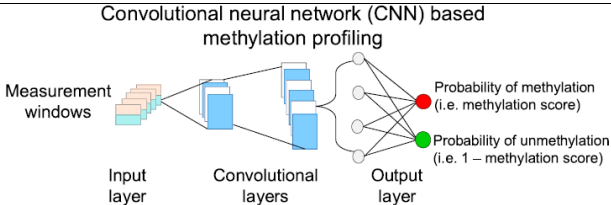
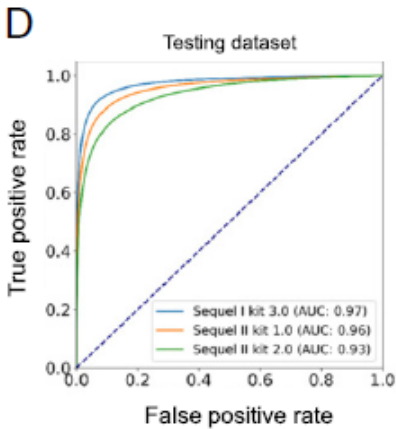
Tse includes a statement that a patent application was filed on the technology described in the article. EX1030, 1. And Tse discloses an embodiment of claim 1 of the’794 patent as set forth the following table. EX2002, ¶60.

Claim 1 of the 794 patent EX1001	Disclosure in Tse (EX1030)
1. A method for detecting a modification of a nucleotide in a nucleic acid molecule, the method comprising:	“Detection of Cytosine Methylation” EX1030, Title.
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,	 Native DNA molecules Single molecule, real-time sequencing Circular DNA template Watson Crick DNA template New strand IPD Pulse strength Time EX1030, Fig. 1 (portion).

Claim 1 of the 794 patent EX1001	Disclosure in Tse (EX1030)
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;	
(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,	The diagram illustrates the process of combining data from Watson and Crick strands. At the top, a DNA template sequence is shown: G A T A C G T T A. Below this, pulse signals are represented as horizontal bars with numerical values. The Watson strand data is shown as a table with columns for each nucleotide (G, A, T, A, C, G, T, T, A) and rows for each nucleotide (G, A, T, C). The Crick strand data is shown as a table with columns for each nucleotide (T, C, A, A, C, G, T, A, G) and rows for each nucleotide (A, C, G, T). The combined data from both strands is shown at the bottom, with columns for each nucleotide (G, A, T, A, C, G, T, T, A) and rows for each nucleotide (G, A, T, C). The combined data is shown as a table with columns for each nucleotide (G, A, T, A, C, G, T, T, A) and rows for each nucleotide (G, A, T, C). EX1030, Fig. 1 (portion).

Claim 1 of the 794 patent EX1001	Disclosure in Tse (EX1030)
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;	
(c) inputting the input data structure into a model, the model trained by: 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	Convolutional neural network (CNN) based methylation profiling  EX1030, Fig. 1 (portion). “We used amplified DNA (unmethylated) and M.SssI-treated DNA (methylated) (M.SssI being a CpG methyltransferase) to train a convolutional neural network.” EX1030 1. “The measurement windows associated with methylated CpG sites (the M.SssI-treated DNA dataset) and those associated with unmethylated CpG sites (the WGA DNA dataset) were used for training the HK model through CNN. Data within each measurement window flanking a cytosine within a CpG context, including the sequence context, mean IPDs, and PWs originating from subreads across individual nucleotides, were entered into the HK model. Each target output (i.e., analogous to a dependent variable value) for a CpG

Claim 1 of the 794 patent EX1001	Disclosure in Tse (EX1030)
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, 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 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,	site in M.SssI-treated DNA datasets was assigned as “1” while each target output for a CpG site in WGA DNA datasets was assigned as “0.” The patterns present in the measurement windows of methylated and unmethylated CpG sites were used for training CNN to determine the parameters (often called weights) of the HK model.” EX1030 10. “The HK model was applicable to data produced by different sequencing kits and sequencers as long as the training and testing processes were based on the same experimental conditions.” EX1030 4.

Claim 1 of the 794 patent EX1001	Disclosure in Tse (EX1030)
(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.	Convolutional neural network (CNN) based methylation profiling  EX1030, Fig. 1 (portion). Also see EX1030, Fig. 2D showing accuracy characterized by AUC to be above 9.3.  D Testing dataset True positive rate False positive rate Sequel I kit 3.0 (AUC: 0.97) Sequel II kit 1.0 (AUC: 0.96) Sequel II kit 2.0 (AUC: 0.93) EX1030, Fig. 2 (portion).

V. CONCLUSION

For the foregoing reasons, Patent Owner requests that the Petition be denied.

Dated: January 26, 2024

Respectfully submitted,

/Michael J. Wise/

Michael J. Wise, Reg. No. 34,047

Lead Counsel

Perkins Coie LLP

1888 Century Park East, Suite 1700

Los Angeles, California 90067-1721

Phone: 310.788.9900

Fax: 310.788.3399

Back-Up Counsel

Joseph P. Hamilton, Reg. No. 51,770

Yingli Wang, Reg. No. 67,372

Attorneys for Patent Owner, The Chinese
University of Hong Kong

CERTIFICATE OF WORD COUNT UNDER 37 CFR § 42.24(d)

Under 37 C.F.R. § 42.24(d), the undersigned certifies that the word count for this Patent Owner's Preliminary Response to the Petition for *inter partes* review totals 12,098, excluding the parts exempted by 37 C.F.R. § 42.24(a). The word count was made using the built-in word count function in the Microsoft® Word software used to prepare this document.

Dated: January 26, 2024

Respectfully submitted,

/Michael J. Wise/

Michael J. Wise, Reg. No. 34,047
Lead Counsel

Perkins Coie LLP
888 Century Park East, Suite 1700
Los Angeles, California 90067-1721
Phone: 310.788.9900
Fax: 310.788.3399

Back-Up Counsel
Joseph P. Hamilton, Reg. No. 51,770
Yingli Wang, Reg. No. 67,372

Attorneys for Patent Owner, The Chinese
University of Hong Kong

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. § 42.6(e), I certify that I caused to be served a true and correct copy of the foregoing: PATENT OWNER'S PRELIMINARY RESPONSE and accompanying exhibits by email to the electronic service addresses for Petitioner on the date indicated below:

LEAD COUNSEL	BACK-UP COUNSEL
Michael T. Rosato (Reg. No. 52,182) mrosato@wsgr.com Wilson Sonsini Goodrich & Rosati, 701 Fifth Avenue, Suite 5100, Seattle, WA 98104-7036 Tel.: 206-883-2529 Fax: 866-947-7329	Sonja R. Gerrard (Reg. No. 72,802) sgerrard@wsgr.com Wilson Sonsini Goodrich & Rosati, 701 Fifth Avenue, Suite 5100, Seattle, WA 98104-7036 Tel.: 206-883-2649 Fax: 866-947-7329
	Patrick M. Medley (Reg. No. 74,735) pmedley@wsgr.com Wilson Sonsini Goodrich & Rosati, 701 Fifth Avenue, Suite 5100, Seattle, WA 98104-7036 Tel.: 206-883-2593 Fax: 866-947-7329

Dated: January 26, 2024

Respectfully submitted,

/Amy Candeloro/

Amy Candeloro, Paralegal
Perkins Coie LLP