

The Patent Trial and Appeal Board
Pacific Biosciences of California, Inc.
Petitioner,
v.
The Chinese University of Hong Kong
Patent Owner

Case No. IPR2024-00028
Patent Owner Demonstratives

'794 Patent Claim 1



US011091794B2

(12) **United States Patent**
Lo et al.

(10) **Patent No.:** **US 11,091,794 B2**
(45) **Date of Patent:** **Aug. 17, 2021**

(54) **DETERMINATION OF BASE
MODIFICATIONS OF NUCLEIC ACIDS**

(71) Applicant: **The Chinese University of Hong
Kong, New Territories (HK)**

2014/0004511 A1 1/2014 Korlach et al.
2016/0017419 A1 1/2016 Chiu et al.
2017/0037464 A1 2/2017 Turner et al.
2017/0159132 A1 6/2017 Okino et al.
2017/0233802 A1 8/2017 Flusberg et al.
2018/0201993 A1 7/2018 Turner et al.
2019/0024162 A1 1/2019 Korlach et al.

(72) Inventors: **Yuk-Ming Dennis Lo**, Hong Kong
(CN); **Rossa Wai Kwun Chiu**, Hong
Kong (CN); **Pelyong Jiang**, Hong Kong
(CN); **Suk Hang Cheng**, Hong Kong
(CN); **Wenlei Peng**, Hong Kong (CN);
On Yee Tse, Hong Kong (CN)

(73) Assignee: **The Chinese University of Hong
Kong, New Territories (HK)**

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

FOREIGN PATENT DOCUMENTS

AU 2012380221 B2 9/2016
CN 104053789 A 9/2014
CN 104053789 B 2/2016
KR 101896167 B1 9/2018
WO 2009013492 A1 1/2009
WO 2010027484 A2 3/2010
WO 2010068289 A2 6/2010
WO 2013170429 A1 11/2013
WO 2013188846 A1 12/2013
WO 2014153757 A1 10/2014
WO 2017012544 A1 1/2017

OTHER PUBLICATIONS

Flusberg et al (Nature Methods, 2010, 7:461-465).
Sugai et al (Clinical Epigenetics, 2017, 9:55, internet pp. 1-10).
Kile et al (PLOS ONE, 2010, 5:e13730, internet pp. 1-9).
Flusberg et al (Nature Methods, 2010, 7:461-465) Supplemental
Data.*
Flusberg, Benjamin A. et al.; "Direct detection of DNA methylation
during single-molecule, real-time sequencing"; Nature Methods;
Jun. 2010, vol. 7, No. 6, pp. 461-465 (7 pages).
Liu, Qian et al.; "Detection of DNA base modifications by deep
recurrent neural network on Oxford Nanopore sequencing data";
Nature Communications; 2019, 10(1):2449; doi: 10.1038/s41467-
019-10168-2, 11 pages.
Ni, Peng et al.; "DeepSignal: detecting DNA methylation state from
Nanopore sequencing reads using deep-learning"; Bioinformatics;
2019, vol. 35, No. 22; doi: 10.1093/bioinformatics/btz276, 10
pages.
Suzuki, Yuta et al.; "Agin: measuring the landscape of CpG methylation
of individual repetitive elements"; Bioinformatics; 2016, vol. 32,
No. 19; pp. 2911-2919.

(Continued)

Primary Examiner — Laura B Goddard

(74) Attorney, Agent, or Firm — Kilpatrick Townsend &
Stockton LLP

(21) Appl. No.: **16/995,607**

(22) Filed: **Aug. 17, 2020**

(65) **Prior Publication Data**
US 2021/0047679 A1 Feb. 18, 2021

Related U.S. Application Data

(60) Provisional application No. 63/051,210, filed on Jul.
13, 2020, provisional application No. 63/019,790,
filed on May 4, 2020, provisional application No.
62/991,891, filed on Mar. 19, 2020, provisional
application No. 62/970,586, filed on Feb. 5, 2020,
provisional application No. 62/887,987, filed on Aug.
16, 2019.

(51) **Int. Cl.**
C12Q 1/6816 (2018.01)
G16B 30/00 (2019.01)
G16B 20/00 (2019.01)
C12N 9/22 (2006.01)
C12Q 1/6851 (2018.01)

(52) **U.S. Cl.**
CPC **C12Q 1/6816** (2013.01); **C12N 9/22**
(2013.01); **C12Q 1/6851** (2013.01); **G16B**
20/00 (2019.02); **G16B 30/00** (2019.02)

(58) **Field of Classification Search**
CPC combination set(s) only.
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,467,976 B2 6/2013 Lo et al.
9,116,118 B2 8/2015 Turner et al.
9,175,338 B2* 11/2015 Flusberg C12Q 1/6869
9,175,341 B2 11/2015 Flusberg et al.
10,438,691 B2 10/2019 Kim et al.
10,465,232 B1 11/2019 Wu et al.
10,468,121 B2 11/2019 Kermani et al.
10,590,484 B2 3/2020 Korlach et al.
2011/0183320 A1 7/2011 Flusberg et al.
2013/0230909 A1 9/2013 Pan et al.

(57) **ABSTRACT**

Systems and methods for using determination of base modification in analyzing nucleic acid molecules and acquiring data for analysis of nucleic acid molecules are described herein. Base modifications may include methylations. Methods to determine base modifications may include using features derived from sequencing. These features may include the pulse width of an optical signal from sequencing bases, the interpulse duration of bases, and the identity of the bases. Machine learning models can be trained to detect the base modifications using these features. The relative modification or methylation levels between haplotypes may indicate a disorder. Modification or methylation statuses may also be used to detect chimeric molecules.

19 Claims, 145 Drawing Sheets
(85 of 145 Drawing Sheet(s) Filed in Color)

Specification includes a Sequence Listing.

Pacific Biosciences of California EXHIBIT - 1001

Pacific Biosciences of California v. The Chinese U. of Hong Kong - IPR2024-00028

'794 Patent Claim 1

What is claimed is:

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

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

for each nucleotide:

- an identity of the nucleotide,
- a position of the nucleotide within the sample nucleic acid molecule,
- a width of the pulse corresponding to the nucleotide, and
- an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(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:

receiving a first plurality of first data structures, each first data structure of the first plurality of data

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

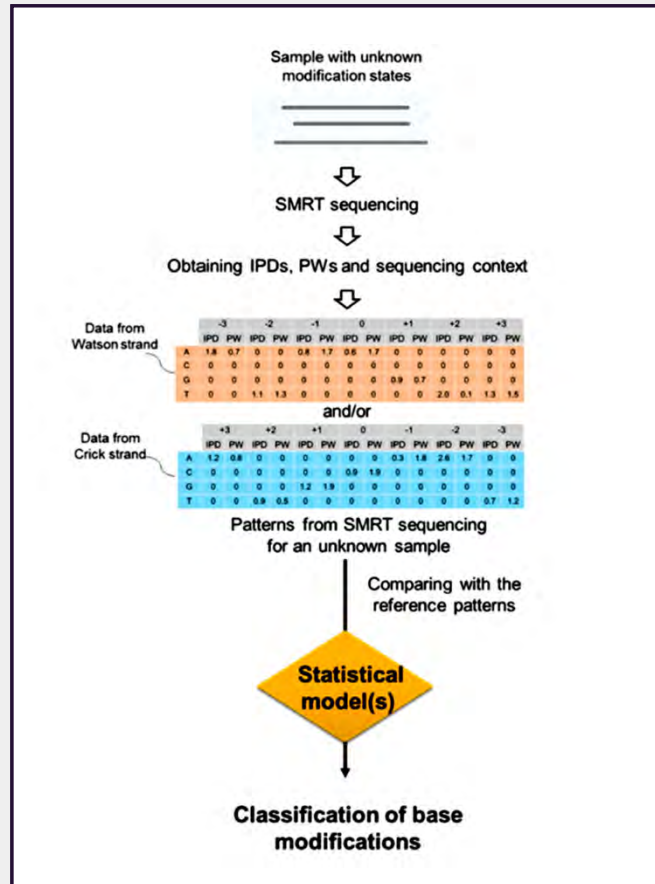
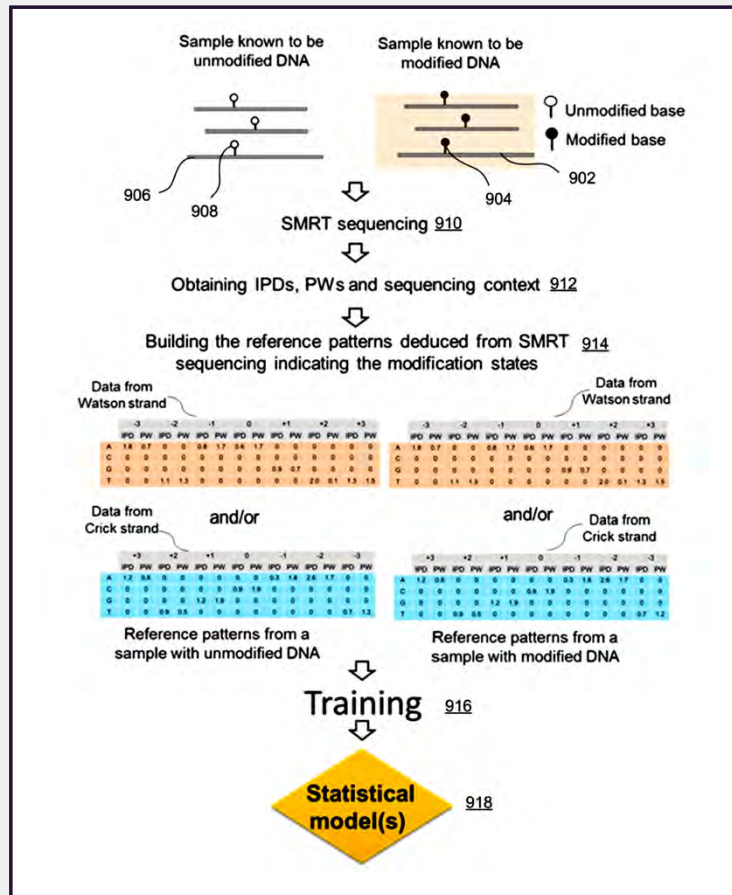
storing a plurality of first training samples, each including one of the first plurality of first data structures and a first label indicating the first state of the nucleotide at the target position, and

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,

(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.

EX1001 ('794 Patent) 115:37-116:61.

Models Are Trained with the Same Input Data Structure Properties as the Input Data Structure



Experts Agree that Model Training Data Must be the Same Format as the Input Data Structure

44. Claim 1 further recites that “each first data structure comprising values for the same properties as the input data structure.” Thus, in accordance with the plain language of claim 1, a POSA would understand that to train the model to detect modifications from an input data structure, the training data must be provided in the same format, *i.e.*, the same window with the same properties (nucleotide identity, nucleotide position relative to a target nucleotide position, nucleotide PW and IPD) as the input data structure that would be used to detect DNA modifications (testing data). I note that Dr. Dessimoz agrees that the input data structure must be so structured.

EX2041 (Koo Decl.) ¶44, p. 17.

130. As I explained for element 1.3, a person of ordinary skill in the art would have understood that to train a model to detect modifications from an input data structure, training data should be provided in the same format—and containing the same properties—as the input data structure that would be used for methylation detection. Thus, the data structures used to train a model that would

EX1002 (Dessimoz Decl.) ¶130, p. 79.

The Flusberg Publications Do Not Teach Including Nucleotide Identity as a Property in an Input Data Structure to Detect Modifications

Basis for Petitioner's Obviousness Case

Relying on the testimony of Dr. Dessimoz, Petitioner argues that a person of ordinary skill in the art “would have found it obvious to consider the claimed properties together because doing so represents nothing beyond the predictable use of known variables according to their well-characterized role in base modification detection.” Pet. 31 (citing Ex. 1002 ¶¶83–84); see *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2017) (“[A] court must ask whether the improvement is more than the predictable use of prior art elements according to their established functions.”).

Paper 8 (Decision Granting Institution) p. 20.

Petitioner's Purported Well-Characterized Role of Sequence Context in SMRT Sequencing Base Modification Detection is it Must be Considered

Because of the documented interdependence of local sequence context and base modifications on PW and IPD, POSAs appreciated that local sequence effects should be considered when predicting base modifications.

Petition, p. 30.

84. Because of the well-documented interdependence of local sequence context and base modifications on PW and IPD, skilled artisans appreciated that local sequence effects had to be considered when predicting base modifications. Flusberg expressly teaches as much, concluding that, “in general, the kinetic signatures of methylation will be sequence context-dependent.” EX1006, 462-463. Other prior art came to the same conclusion; for example, Schadt et al. found that sequence context very strongly affected polymerase kinetics, noting that the context over a window running “seven bases upstream of and two bases downstream from the incorporation site explain >80% of the variance in the IPD measures.” EX1010, 130; *see also supra* Sections V.B-C. A person of ordinary skill in the art would have appreciated that, because these different properties were interrelated, they should be considered together when classifying modifications across varying sequence contexts. For example, a person of ordinary skill in the art would understand that, based on the prior art’s recognition that sequence context accounts for >80% of the variance in IPD, any variation in that context must be accounted for when detecting modifications using kinetic parameters like IPD, or else the sequence context will act as an uncontrolled confounder that drowns the signal for methylation detection in noise. A person of ordinary skill in the art

EX1002 (Dessimoz Decl.) ¶184, p. 45.

Petitioner's Expert Asserts that Sequence Context Must be Controlled for or Explicitly Evaluated

Because the sequence context (e.g., identity of the bases in the vicinity of a modified base) strongly affects IPD and PW, it was further appreciated that context must be controlled for or explicitly evaluated when predicting base modifications.

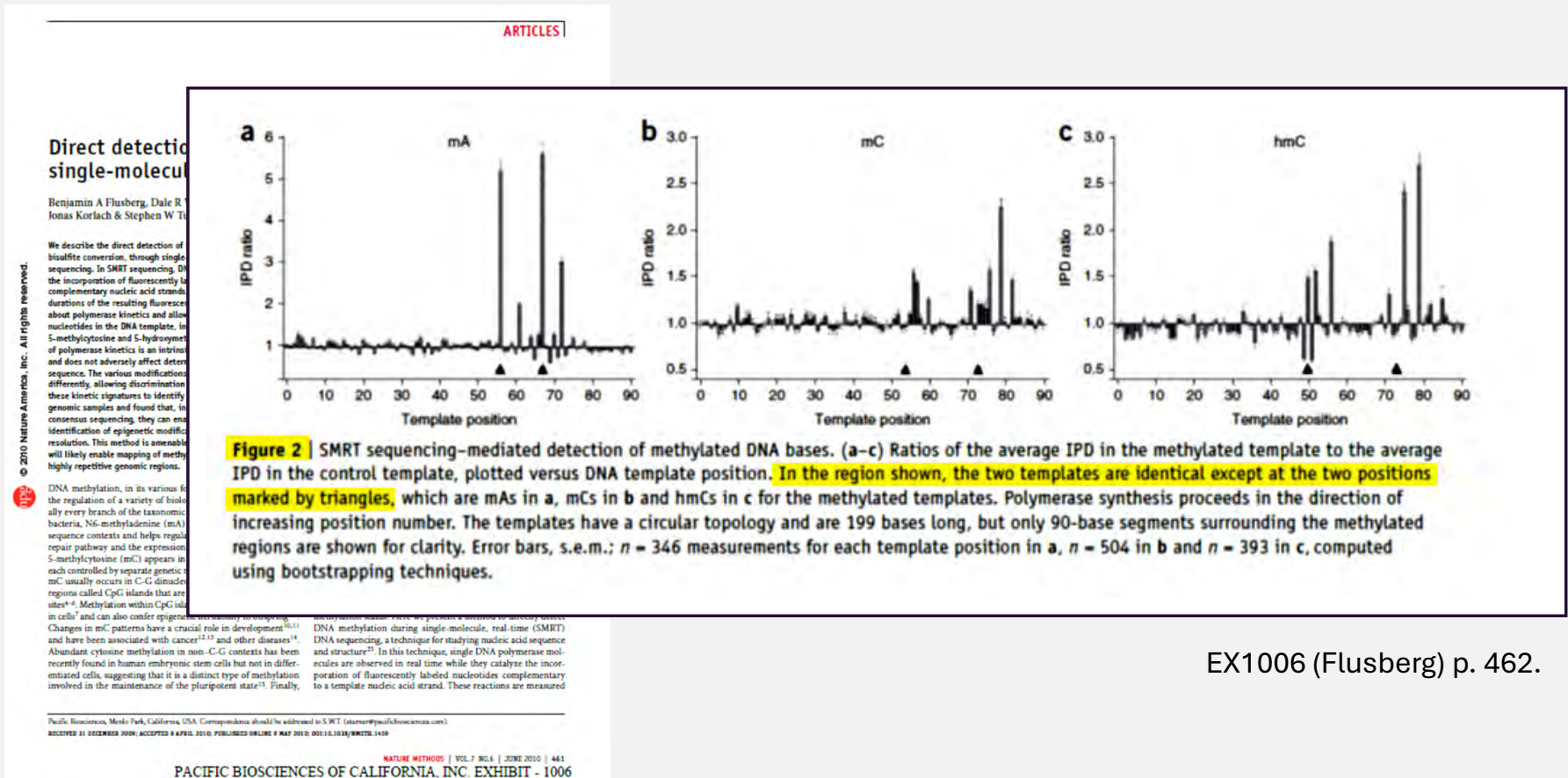
EX1002 (Dessimoz Decl.) ¶35, p. 21.

Flusberg Controls for Sequence Context

and pulse width. 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 nucleotide sequence. See EX1006, 465 (showing same nucleotide sequences used for Figs. 1-4, with only modification status varied). Thus, the

EX1002 (Dessimoz Decl.) ¶107, p. 63.

“Flusberg’s specific example avoids using sequence context as an express input variable because it is held constant” EX1002 (Dessimoz Decl.) ¶107.



EX1006 (Flusberg) p. 462.

“showing same nucleotide sequences used for Figs. 1-4, with only modification status varied” EX1002 (Dessimoz Decl.) ¶107.

Direct detection of DNA methylation during single-molecule, real-time sequencing

Benjamin A Flusberg, Dale R Webster, Jessica H Lee, Kevin J Travers, Eric C Olivares, Tyson A Clark, Jonas Korlach & Stephen W Turner

We describe the direct detection of DNA methylation, without bisulfite conversion, through single-molecule, real-time (SMRT) sequencing. In SMRT sequencing, DNA polymerases catalyze the incorporation of fluorescently labeled nucleotides into complementary nucleic acid strands. The arrival times and durations of the resulting fluorescence pulses yield information about polymerase kinetics and allow direct detection of modified nucleotides in the DNA template, including N6-methyladenine, 5-methylcytosine and 5-hydroxymethylcytosine. Measurement of polymerase kinetics is an intrinsic part of SMRT sequencing and does not adversely affect determination of primary DNA sequence. The various modifications affect polymerase kinetics differently, allowing discrimination between them. We used these kinetic signatures to identify adenine methylation in genomic samples and found that, in combination with circular consensus sequencing, they can enable single-molecule identification of epigenetic modifications with base-pair resolution. This method is amenable to long read lengths and will likely enable mapping of methylation patterns in even highly repetitive genomic regions.

DNA methylation, in its various forms, has been implicated in the regulation of a variety of biological processes across virtually every branch of the taxonomic tree. For example, in certain bacteria, N6-methyladenine (mA) appears primarily in GATC sequence contexts and helps regulate replication, the mismatch repair pathway and the expression of certain genes¹. In plants, 5-methylcytosine (mC) appears in multiple sequence contexts, each controlled by separate genetic mechanisms^{2,3}. In vertebrates, mC usually occurs in C-G dinucleotides, which often cluster in regions called CpG islands that are at or near transcription start sites⁴⁻⁶. Methylation within CpG islands regulates gene expression in cells⁷ and can also confer epigenetic heritability in offspring^{8,9}. Changes in mC patterns have a crucial role in development^{10,11} and have been associated with cancer^{12,13} and other diseases¹⁴. Abundant cytosine methylation in non-C-G contexts has been recently found in human embryonic stem cells but not in differentiated cells, suggesting that it is a distinct type of methylation involved in the maintenance of the pluripotent state¹⁵. Finally,

5-hydroxymethylcytosine (hmC) is a newly identified epigenetic mark whose biological function is not yet understood, found thus far in mouse Purkinje neurons¹⁶ and embryonic stem cells¹⁷.

Because of its key role in human health and disease, cytosine methylation is the most widely studied of the DNA modifications described above, and there is much interest in mapping genome-wide mC patterns across different cell types and in response to various environmental influences¹⁸. Currently, the most common technique for studying cytosine methylation involves bisulfite treatment (which transforms epigenetic information to genetic information by converting cytosine, but not methylcytosine, to uracil) followed by massively parallel DNA sequencing¹⁹. Using this approach, researchers have recently constructed single-base resolution methylation maps for the *Arabidopsis thaliana* genome^{2,3}, for a subset of the mouse genome¹⁹ and for both fibroblasts and embryonic stem cells throughout the majority of the human genome²⁰. Despite these advances, enabled by bisulfite treatment-based sequencing, there are several drawbacks to the technique. For example, the sample preparation steps associated with bisulfite treatment can be costly and time-consuming, and the harsh reaction conditions necessary for complete conversion can degrade DNA. Additionally, the reduction of complexity in converted genomes constrains primer design for subsequent PCR amplification²⁰ and complicates alignment to a reference genome⁶. Finally, discrimination between cytosine, mC and hmC cannot be accomplished with bisulfite treatment-based sequencing^{16,17,21,22}.

Direct detection of methylation is possible for nucleotides in solution using techniques such as thin layer chromatography^{16,17}, high-performance liquid chromatography^{16,23}, mass spectrometry^{16,17,23} and nanopore amperometry²⁴. To our knowledge, however, no high-throughput method has been demonstrated that allows the determination of primary sequence at the same time as methylation status. Here we present a method to directly detect DNA methylation during single-molecule, real-time (SMRT) DNA sequencing, a technique for studying nucleic acid sequence and structure²⁵. In this technique, single DNA polymerase molecules are observed in real time while they catalyze the incorporation of fluorescently labeled nucleotides complementary to a template nucleic acid strand. These reactions are measured

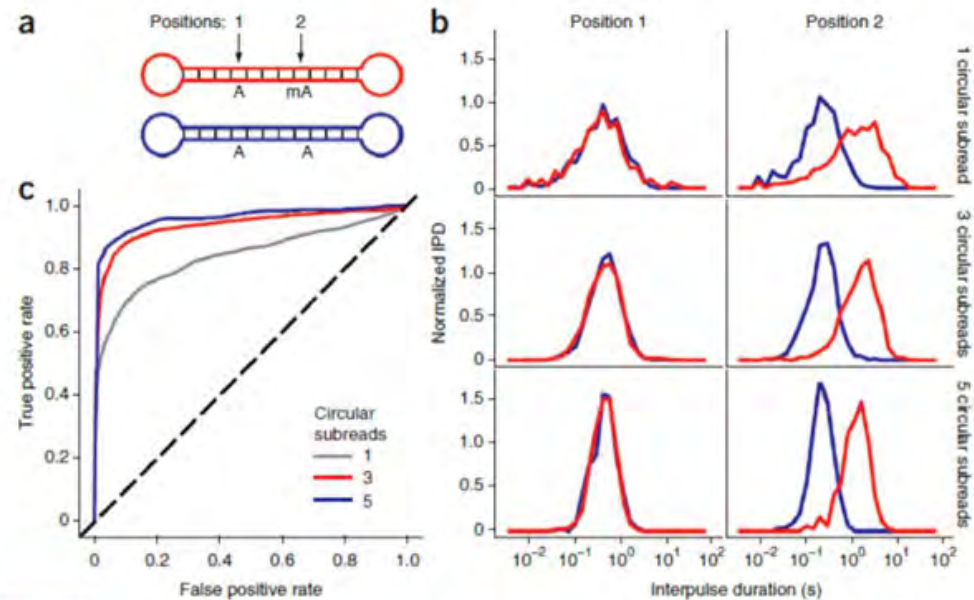


Figure 4 | IPD distributions for adenine and mA in synthetic DNA templates. (a) Schematic of the 199-base DNA templates used in this experiment. (b) IPD distributions at the indicated positions for both templates, colored as in a. The histograms depict the distributions of mean IPD (averaged over the indicated number of circular subreads). (c) Receiver operating characteristic curves, based on the IPD distributions from the differentially methylated position in b and parameterized by IPD threshold, for assigning a methylation status to an adenine nucleotide after 1, 3 and 5 circular consensus sequencing subreads. The dashed line depicts the receiver operating characteristic curve for randomly guessing the methylation status.

Pacific Biosciences, Menlo Park, California, USA. Correspondence should be addressed to S.W.T. (sturner@pacificbiosciences.com).
RECEIVED 31 DECEMBER 2009; ACCEPTED 9 APRIL 2010; PUBLISHED ONLINE 9 MAY 2010; DOI:10.1038/NMETH.1408

NATURE METHODS | VOL. 7 NO. 6 | JUNE 2010 | 441

PACIFIC BIOSCIENCES OF CALIFORNIA, INC. EXHIBIT - 1006

Both Flusberg References Control for Sequence Context

19 In this case, we cannot fix the pulse
20 width or the interpulse duration. These are just
21 observations. But we can fix the sequence context.
22 And I think this is what they did in this -- in this
23 experiment reported in the Flusberg publications.

24 BY MR. WISE:

Q And it's true in the patent, too, right?

1 In the examples in the patent, they're
2 using unmodified templates identical to the modified
3 templates, and then measuring the different -- the
4 kinetic differences, right?

5 MS. GERRARD: Objection. Form.

6 THE WITNESS: By and large, there's a big
7 overlap between the patent and the publication.
8 This is why I refer to them jointly. And, yes, I
9 mean, as far as these general statements go, what is
10 true for one is likely true for the other.

EX2037 (Dessimoz Dep.) 87:19-88:10.

17 Q That's what I'm trying to understand.
18 So we talk about Flusberg which controls
19 for sequence context, right? That's what Flusberg
20 is doing, agreed?
21 A Yes.

EX2037 (Dessimoz Dep.) 90:17-21.

“By and large, there's a big overlap between the patent and the publication.” EX2037 (Dessimoz Dep.) 87:19-88:10.

United States Patent
Flusberg et al.

(11) Patent No.: **US 9,175,338 B2**
(45) Date of Patent: ***Nov. 3, 2015**

(54) **METHODS FOR IDENTIFYING NUCLEIC ACID MODIFICATIONS**

(75) Inventors: **Benjamin Flusberg**, Atlanta, GA (US); **Jeann Korfback**, Newark, CA (US); **Jeffrey Wegener**, Cupertino, CA (US); **Iyoon A. C. Park**, Menlo Park, CA (US); **Igor Vilfan**, East Palo Alto, CA (US); **Andrey Kichuk**, San Ramon, CA (US); **Stephen Turner**, Menlo Park, CA (US); **Jon Sorenson**, Alameda, CA (US); **Kevin Travers**, Menlo Park, CA (US); **Cheryl Heiner**, San Mateo, CA (US); **Austin B. Tomasey**, Burlingame, CA (US); **Patrick Marks**, San Francisco, CA (US); **Jessica Lee**, Cupertino, CA (US); **Lei Jia**, Palo Alto, CA (US); **Dale Webster**, San Mateo, CA (US); **John Lyke**, Redwood Shores, CA (US); **Jeremiah Hanon**, Redwood City, CA (US); **Joseph Puglisi**, Stanford, CA (US)

(79) Assignee: **Pacific Biosciences of California, Inc.**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days. This patent is subject to a terminal disclaimer.

(21) Appl. No.: **12/948,767**

(22) Filed: **Nov. 12, 2010**

(65) **Prior Publication Data**
US 2011/0183320 A1, filed Nov. 28, 2011

Related U.S. Application Data
(63) Continuation-in-part of application No. 12/635,618, filed on Dec. 16, 2009.
(60) Provisional application No. 61/201,551, filed on Dec. 11, 2009; provisional application No. 61/180,550, filed on May 21, 2009; provisional application No. 61/186,661, filed on Jun. 12, 2009.

(51) Int. Cl. **C72Q 1/68** (2006.01)

(52) U.S. Cl. **C72Q 1/68** (2013.01); **C72Q 1/68** (2013.01); **C72Q 1/68** (2013.01)

(58) **Field of Classification Search**

See application file for complete search history.

References Cited

U.S. PATENT DOCUMENTS

5,347,838 A 8/1998 Dwyer et al.
6,792,806 B1 4/2001 Chae
8,235,502 B1 5/2001 Watanabe et al.
8,255,083 B1 7/2001 Williams

15 Claims, 28 Drawing Sheets

PACIFIC BIOSCIENCES OF CALIFORNIA, INC. EXHIBIT - 1007

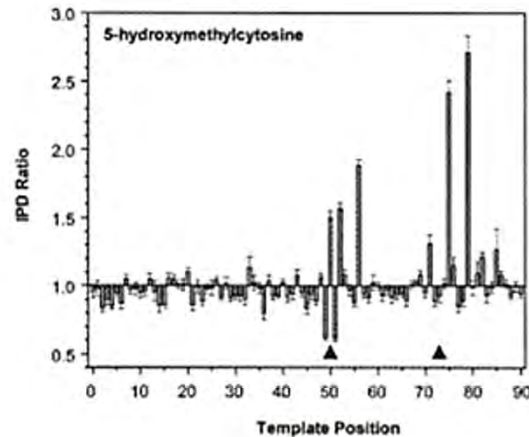


Figure 16

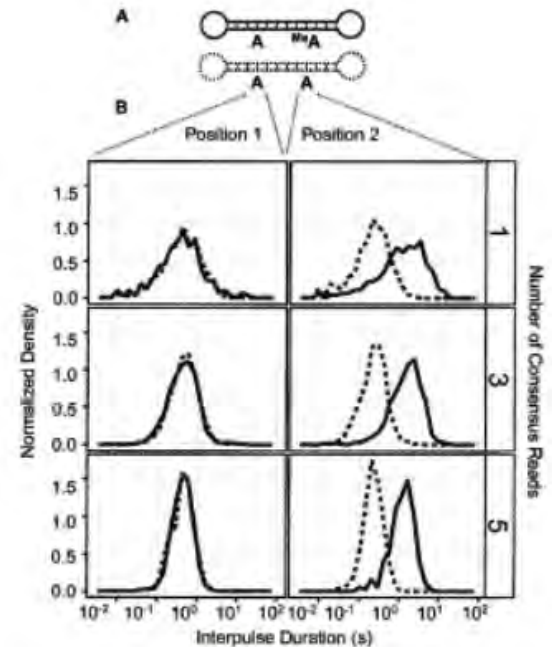


Figure 14

The Flusberg '338 Patent Does Not Disclose Including the Identity of Each Nucleotide as a Property of an Input Data Structure

entirety for all purposes. 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

$$\vec{f} = \{f_1^{i-3}, f_2^{i-3}, \dots, f_k^{i-3}, f_1^{i-2}, f_2^{i-2}, \dots, f_k^{i-2}, f_1^{i-1}, f_2^{i-1}, \dots, f_k^{i-1}, f_1^i, f_2^i, \dots, f_k^i, f_1^{i+1}, f_2^{i+1}, \dots, f_k^{i+1}, f_1^{i+2}, f_2^{i+2}, \dots, f_k^{i+2}, f_1^{i+3}, f_2^{i+3}, \dots, f_k^{i+3}\}.$$

EX1007 ('338 Patent) 66:54-67.

tions, as well. Further, in light of the above teachings it will be clear to one of ordinary skill that the approach can be extended to pulse metrics other than IPD, such as pulse width, branch rate, mismatch rate, deletion rate, etc. In addition, the

EX1007 ('338 Patent) 67:64-67.

The PacBio Art Controls for Sequence Context

135. To support his opinion that a person of ordinary skill in the art would have “good reason to combine the Flusberg publications with PacBio-BAM to achieve what is recited in the claims with a reasonable expectation of success,” Dr. Dessimoz cites several papers from PacBio researchers and collaborators: Flusberg (2010), EX 1006, the ’338 patent, EX1007, Clark (2011), EX1008, Loomis (2013), EX1009, Schadt (2013), EX1010, and PacBio’s White Paper EX1016. But, none of those papers discloses including the identity of a nucleotide in an input data structure for base modification as claimed. To the contrary each of those references either follows the proposed methods of the Flusberg publications that control for nucleotide identity by making it constant, or does not include any method of detection [of] a nucleotide modification.

The White Paper (Ex1016) and kineticsTools (Ex1021) Control for Sequence Context

2 Q I think if we look at the last two
3 sentences, it kind of makes it clear. "An amplified
4 control dataset is generated by sequencing
5 unmodified DNA with the same sequence as the test
6 sample."

7 Would you agree that this is fixed just
8 like the white paper?

9 MS. GERRARD: Objection. Form.

10 THE WITNESS: Well, as the name indicates,
11 it is controlled in this case, and, yes, I think
12 this is consistent with the method that is described
13 in the white paper or in the Flusberg publications.

The Well-Known Use of Nucleotide Identity Was to Keep it Constant

in 2019. To the contrary, the well-known use of nucleotide identity in detecting base modification with either PW and/or IPD at the time was to control for nucleotide identity by making it constant. That a POSA might have understood that sequence context or even nucleotide identity should be considered in nucleotide modification detection does not suggest to a POSA which properties related to sequence context should be considered or how to account for any one of those properties including how to use nucleotide identity.

EX2041 (Koo Decl.) ¶149, pp. 69-70.

The PacBio-BAM Does Not Teach Including Nucleotide Identity as a Property in an Input Data Structure to Detect Modifications

The Experts Agree that PacBio-BAM Does Not Teach Modification Detection

Perkins
Coie

explained above. The PacBio-BAM file is the stored data generated in the normal course of SMRT sequencing and does not teach or suggest any particular method of modification detection. A POSA would not be motivated to modify the methods of the Flusberg publications that control for nucleotide identity, and instead include nucleotide identity as a property in the claimed input data structure.

EX2041 (Koo Decl.) ¶132, p. 61.

20 A That's right, but I would like -- I do not
21 want to be misunderstood here.
22 I do not claim that this document teaches
23 how to do base modification detection. This is
24 merely a description of the content of this file.

EX2037 (Dessimoz Dep.) 105:20-24.

10 But to the extent -- but it is not a
11 document that provides much information in terms of
12 base modification detection.

EX2037 (Dessimoz Dep.) 106:10-12.

PacBio BAM Saves Disk Space for Kinetic Features

Encoding of kinetics pulse features

Interpulse duration (IPD) and pulsewidth are measured in frames; natively they are recorded as a `uint16` per pulse/base event. They may be encoded in BAM read tags in one of two fashions:

- losslessly as an array of `uint16`; necessary for PacBio-internal applications but entails greater disk space usage.
- lossy 8-bit compression stored as a `uint8` array, following the codec specified below ("codec V1").
Provides a substantial disk-space savings without affecting important production use cases (base modification detection).

In the default production instrument configuration, the lossy encoding will be used. The instrument can be switched into a mode (PacBio-internal mode) where it will emit the full lossless kinetic features.

EX1011 (PacBio BAM) p. 8.

The PacBio-BAM Does Not Teach Modification Detection

131. The PacBio BAM's reference that kinetic data can be used for "base modification detection" does not teach or suggest to a person of ordinary skill in the art how to create an input data structure or to include the combination of features of the claimed input data structure. EX1011 at 8. Nucleotide identity is not a kinetic pulse feature. And, the PacBio BAM does not disclose or suggest any particular method for base modification detection. Instead, the PacBio-BAM indicates that its format saves disk space without affecting the kinetic data (IPD and PW) for use cases like "base modification detection." *Id.* Based on the disclosure in the PacBio-BAM, a POSA would not be motivated to modify the methods of the Flusberg publications, that control for nucleotide identity by keeping it constant, and instead include nucleotide identity as a property in an input data structure as claimed.

No Reason to Modify Flusberg's
Method of Controlling for
Nucleotide Identity to Include it as
an Input Data Structure Property

No Motivation to Modify a Method that Avoids Using Nucleotide Identity as a Property in an Input Data Structure

132. Flusberg and the '338 patent do not teach or suggest the use of nucleotide identity as a property in an “input data structure” as claimed as I have explained above. The PacBio-BAM file is the stored data generated in the normal course of SMRT sequencing and does not teach or suggest any particular method of modification detection. A POSA would not be motivated to modify the methods of the Flusberg publications that control for nucleotide identity, and instead include nucleotide identity as a property in the claimed input data structure.

EX2041 (Koo Decl.) ¶132, p. 61.

Including Nucleotide Identity as a Property of an Input Data Structure is Not Interchangeable with Keeping it Constant

150. Moreover, in 2019 a POSA would not understand or have reason to believe that the method of the Flusberg publications, and the several references that follow that same approach as Flusberg as explained above, that control for nucleotide identity by making it constant would be interchangeable with the claimed method that includes nucleotide identity as a property in an input data structure as recited in claim 1.

EX2041 (Koo Decl.) ¶150, p. 70.

Including Nucleotide Identity as a Property of an Input Data Structure is Not Interchangeable with Keeping it Constant

Perkins
Coie

20 Q Okay. Previously, we were talking earlier
21 today, and you testified that sequence context needs
22 to be taken into account, and we discussed two ways
23 of doing it: The Flusberg method and the '794
24 method.

25 Do you recall that?

1 MS. GERRARD: Objection. Form.

2 THE WITNESS: I believe I know what you're
3 referring to. Yes.

4 BY MR. WISE:

5 Q Would a person of skill in the art
6 consider those two methods of taking into account
7 sequence context interchangeable?

8 MS. GERRARD: Objection. Form.

9 THE WITNESS: I mean, these are different
10 kind of experiments, and the -- so, no, they cannot
11 be viewed interchangeably. It's quite a different
12 situation.

EX2037 (Dessimoz Dep.) 108:20-109:12.

The Flusberg Method and the Method of Claim 1 Are Not Comparable

14 Q Would a person of skill in the art expect
15 similar results between those two different
16 experiments?

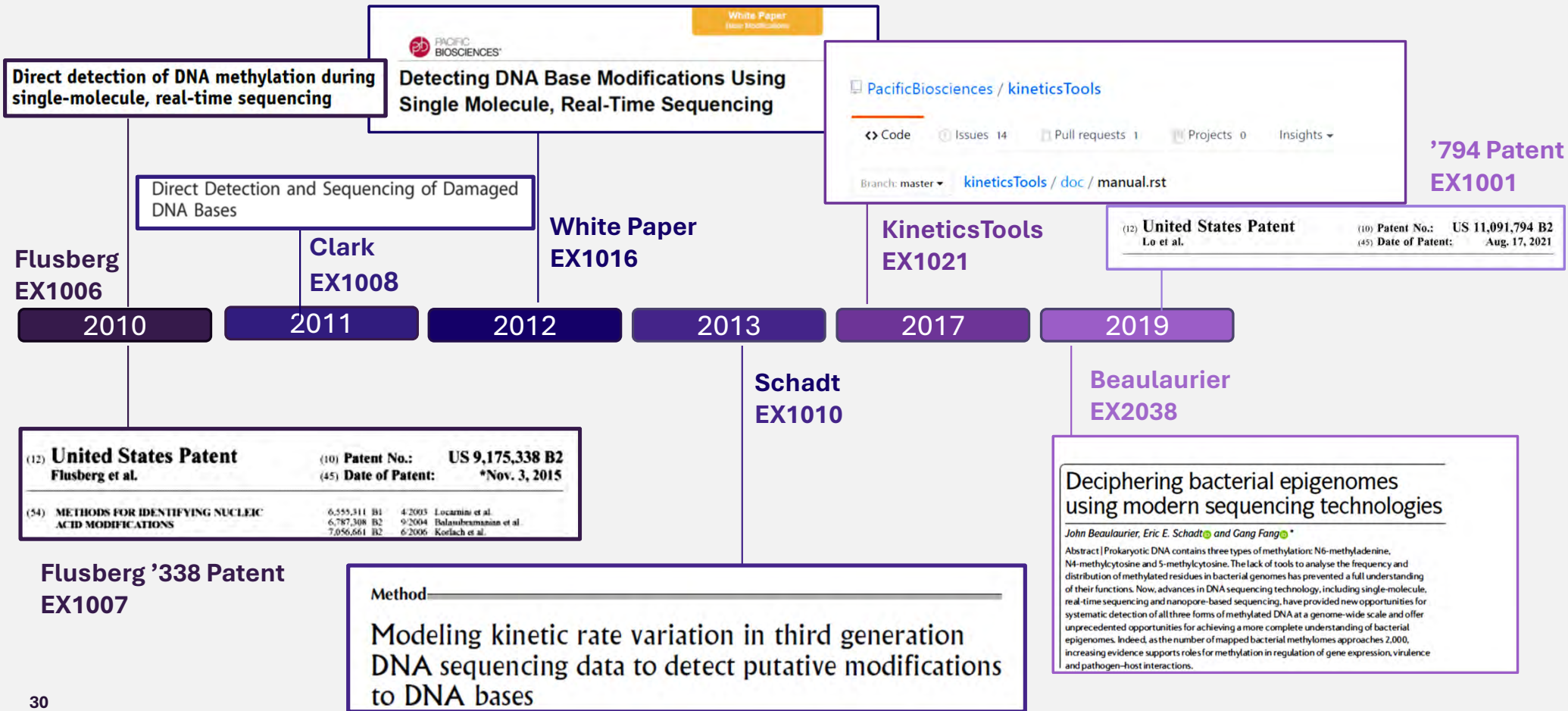
17 MS. GERRARD: Objection. Foundation.

18 THE WITNESS: So if taken as a whole, are
19 you asking me about whether the results described
20 in -- in the Flusberg publication, whether we might
21 expect the same results in the '794 experiments?

22 Well, then my answer would be no, because
23 there are many differences between -- between the
24 two.

There Was No Well Characterized
Use of PW in Base Modification
Detection at the Time of the
Invention

PacBio Taught the Art to Use IPD Not PW for Modification Detection



The White Paper Teaches to Use IPD for Modification Detection

White Paper
Base Modifications

PACIFIC BIOSCIENCES

Detecting DNA Base Modifications Using Single Molecule, Real-Time Sequencing

Introduction

Base modification understanding expression, damage, and Time (SMRT) revolutionizing direct detection of unamplified source material.

Traditionally, it has been a challenge to study the wide variety of base modifications that are seen in nature. Most high-throughput techniques focus only on cytosine methylation and involve both bisulfite sequencing to convert unmethylated cytosine nucleotides to uracil nucleotides, and comparison of sequence reads from bisulfite-treated and untreated samples². SMRT sequencing, in contrast, does not require base conversion within the source material in order to detect base modifications. Instead, the kinetics of base addition is measured during the normal course of sequencing. These kinetic measurements present characteristic patterns in response to a wide variety of base modifications – researchers at PacBio have been able to observe unique kinetic characteristics for over 25 base modifications³.

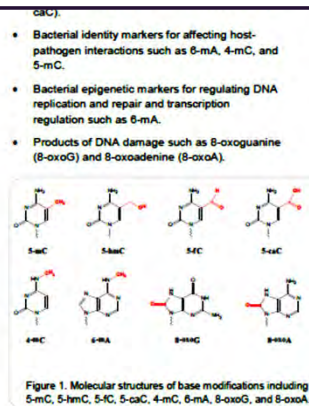
As a result of this breakthrough methodology to detect base modifications, it is now possible to sequence modifications other than 5-methylcytosine (5-mC). Bacterial modifications such as 6-methyladenine (6-mA), 4-methylcytosine (4-mC), or more recently identified eukaryotic modifications such as 5-hydroxymethylcytosine⁴ (5-hmC), are also accessible to study using a single sequencing method on the PacBio[®] RS system.

As our understanding of kinetic information grows, the analysis of base modifications using SMRT technology will continue to become easier and faster.

www.pacb.com/base-mod

PACIFIC BIOSCIENCES OF CALIFORNIA, INC. EXHIBIT - 1016

Figure 2. Principle of detecting modified DNA bases during SMRT sequencing. The presence of the modified base in the DNA template (top), shown here for 6-mA, results in a delayed incorporation of the corresponding T nucleotide, i.e. longer interpulse duration (IPD), compared to a control DNA template lacking the modification (bottom).³



Today, the most commonly studied base modification is 5-mC, commonly referred to as "methylation" despite the fact that other bases are also naturally modified by methyl groups.

EX1016 (White Paper), p. 2.

structure" as recited in claim 1. EX1016 at 3. **PacBio's White Paper (EX1016)** discloses using IPD, but does not disclose using PW. PacBio's White Paper describes two techniques for detecting DNA modifications both using IPD as the input: A comparison of IPD ratio between a DNA sample that is divided into native

EX2041 (Koo Decl.) ¶142, p. 66.

Schadt Teaches to Use IPD for Modification Detection

Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases

Eric E. Schadt,^{1,8,9} Onureena Banerjee,^{2,8} Gang Fang,^{3,8} Zhixing Feng,¹ Wing H. Wong,⁴ Xuegong Zhang,⁵ Andrey Kislyuk,² Tyson A. Clark,¹ Alona Keren-Paz,¹ Andrew Chess,¹ Vipin Kumar,⁶ Alice Chen-Plotkin,¹ Neal Sondheimer,⁷ Jonas Korlach,² and Andrew Kasarskis¹

¹Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York; ²Genetics, Menlo Park, California 94025, USA; ³Department of Computer Science and Engineering, U Minnesota, Minneapolis, Minnesota 55455, USA; ⁴Department of Statistics, Stanford University, Stanford, California; ⁵National Laboratory for Information Science and Technology, and Department of Automation, Tsinghua China; ⁶Department of Neurology, ⁷Department of Pediatrics, The University of Pennsylvania, Philadelphia, Pennsylvania; ⁸Department of Biostatistics, ⁹Department of Epidemiology, The University of Pennsylvania, Philadelphia, Pennsylvania

Current generation DNA sequencing instruments are moving closer to seamlessly sequencing applications as a routine part of scientific investigation. However, while significant inroads have been made in nucleotide variation and structural variations in DNA that impact phenotypes of interest, progress regarding epigenetic changes and base-level damage to DNA, largely due to technological limitations, has been slow. Recently, single-molecule real time (SMRT) sequencing has been reported to identify kinetic variation (KV) events that have been demonstrated to reflect every known type, providing a path forward for detecting base modifications as a routine part of sequencing. Here, we propose a statistical framework to enhance the power to detect these events with false-positive events. By modeling enzyme kinetics in the neighborhood of an arbitrary location of interest as a conditional random field, we provide a statistical framework for incorporating a test position of interest as well as at neighboring sites that help enhance the power to detect modifications. The performance of this and related models is explored, with the best-performing model applied to from *Escherichia coli* and mitochondrial DNA isolated from human brain tissue. We highlight modification events, some of which strongly associate with known modification events, while other chemically modified sites of unknown types.

[Supplemental material is available for this article.]

DNA sequencing carried out using current first and second (or next) generation sequencing (NGS) instruments has achieved significant success in providing DNA sequences in which the A's, G's, C's, and T's comprising any given DNA template of interest are very consistently called at accuracies that exceed 99%, enabling an explosion of whole-genome sequencing applications that promise to transform our understanding of living systems and usher in the era of personalized genomics (Shendure and Ji 2008; Wheeler et al. 2008; Huang et al. 2009; Wall et al. 2009; Alexander et al. 2010; Drmanac et al. 2010; Flusberg et al. 2010; Kan et al. 2010; Qin et al. 2010; van Bakel et al. 2010). However, extensive characterization of the chemical composition of bases encountered in DNA sequences has resulted in the enumeration of many different types of bases existing in nature that go beyond the fundamental deoxyribonucleotides (A's, G's, C's, and T's) that NGS technologies

produce as the end product of their chemical modifications and damage events (5-mC, 5-hydroxymethylcytosine (5-hmC), 8-oxoguanine (8-oxoG), and other nucleotides, have been shown to play a role in the regulation of genes and genomes. Presently, only 5-mC can be routinely detected by NGS technology and then only after sample preparation that creates a number of challenges (Cokus et al. 2010).

Single-molecule real-time (SMRT) technology is the first commercially available platform to directly observe single molecules of DNA polymerase as they synthesize DNA sequences, not only leveraging the speed and processivity of this enzyme to address many of the shortcomings of second-generation sequencing technologies (Levine et al. 2003; Eid et al. 2009), but providing a time dimension from which the kinetics of the DNA polymerase enzyme can be inferred in real time as DNA sequencing is carried out (Schadt et al. 2010). That is, as the DNA polymerase is synthesizing

the interpulse duration or IPD). While there is a fair degree of stochastic variation in the IPD and PW values over different reads covering a common genomic region, reproducible variations in IPD and PW are induced as a result of modifications to the bases, as has been previously shown (Flusberg et al. 2010). 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 (Schadt) pp. 129, 130.

*These authors contributed equally to this work.

*Corresponding author

E-mail: eric.schadt@mssm.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gj.136799.111>.

Freely available online through the Genome Research Open Access option.

23.129-141 © 2013, Published by Cold Spring Harbor Laboratory Press, ISSN 1088-9051/13; www.genome.org

Genome Research 129
www.genome.org

kineticsTools Teaches to Use IPD for Modification Detection

The Wayback Machine - <http://web.archive.org/web/20170813085437/https://github.com/PacificBiosciences/kineticsTools>

Features Business Explore Marketplace Pricing this repository

PacificBiosciences / kineticsTools

Code Issues 14 Pull requests 1 Projects 0 Insights

Branch: master kineticsTools / doc / manual.rst

hatchale note about pbcore/pipeline pipelines
1 contributor

227 lines (189 sloc) 33.6 KB

Overview

kineticsTools loads IPDs observed at each position in the genome, and compares those IPDs to value expected for unmodified DNA, and outputs the result of this statistical test. The expected IPD value for unmodified DNA can come from either an *in-silico* control or an *amplified control*. The *in-silico* control is trained by PacBio and shipped with the package. It predicts the IPD using the local sequence context around the current position. An amplified control dataset is generated by sequencing unmodified DNA with the same sequence as the test sample. An amplified control sample is usually generated by whole-genome amplification of the original sample.

Dependencies

kineticsTools depends on:

- pbcore (<http://github.com/PacificBiosciences/pbcore>)
- h5py
- numpy
- scipy

Modification Detection

The basic mode of kineticsTools does an independent comparison of IPDs at each position on the genome, for each strand, and emits various statistics to CSV and GFF (after applying a significance filter).

Modifications Identification

kineticsTools also has a *Modification Identification mode* that can decode multi-site IPD 'fingerprints' into a reduced set of calls of specific modifications. This feature has the following benefits:

- Different modifications occurring on the same base can be distinguished (for example m5C and m4C)
- The signal from one modification is combined into one statistic, improving sensitivity, removing extra peaks, and correctly centering the call

Algorithm

Synthetic Control

Studies of the relationship between IPD and sequence context reveal that most of the variation in mean IPD across a genome can be predicted from a 12-base sequence context surrounding the active site of the DNA polymerase. The bounds of the relevant context window correspond to the window of DNA in contact with the polymerase, as seen in DNA/polymerase

Pg. 1
PACIFIC BIOSCIENCES OF CALIFORNIA, INC. EXHIBIT - 1021

Overview

kineticsTools loads IPDs observed at each position in the genome, and compares those IPDs to value expected for unmodified DNA, and outputs the result of this statistical test. The expected IPD value for unmodified DNA can come from either an *in-silico* control or an *amplified control*. The *in-silico* control is trained by PacBio and shipped with the package. It predicts the IPD using the local sequence context around the current position. An amplified control dataset is generated by sequencing unmodified DNA with the same sequence as the test sample. An amplified control sample is usually generated by whole-genome amplification of the original sample.

EX1021 (kineticsTools Manual) p. 1.

15 Q Okay. Does kineticsTools use any other
16 kinetic data other than IPD?
17 A I would need to review this briefly. I
18 don't remember the content.
19 Q Please do.
20 A So based on my reading, it does not appear
21 to be using pulse width.

EX2037 (Dessimoz Dep.) 186:15-21.

By 2019 PacBio Still Detected Modifications with IPD

Deciphering bacterial epigenomes using modern sequencing technologies

John Beaulaurier, Eric E. Schadt^{ID} and Gang Fang^{ID}*

N4-methylcytosine and 5-methylcytosine. The lack of tools to analyse the frequency and distribution of methylated residues in bacterial genomes has prevented a full understanding of their functions. Now, advances in DNA sequencing technology, including single-molecule real-time sequencing and nanopore-based sequencing, have provided new opportunities for systematic detection of all three forms of methylated DNA at a genome-wide scale and unprecedented opportunities for achieving a more complete understanding of bacterial epigenomes. Indeed, as the number of mapped bacterial methylomes approaches 2, increasing evidence supports roles for methylation in regulation of gene expression, and pathogen–host interactions.

Phase variation
A means by which reversible variation of protein expression is achieved in bacteria, often in an ON/OFF manner. The process creates phenotypic diversity in a clonally expanded population and allows the colony to survive in rapidly changing environments.

Adaptive selection
An evolutionary process through which surviving organisms accumulate genetic changes that lead to a fitness advantage over their progenitors.

DNA methylation was discovered in bacteria more than a half century ago¹. It is now known that modification of the four canonical DNA bases by methylation can act as an epigenetic regulator — that is, it can impart distinct and reversible regulatory states to identical genetic sequences. In eukaryotes, epigenetic regulation can occur at multiple levels: DNA methylation, nucleosome positioning, histone variants and histone modifications. By contrast, bacteria lack histones and nucleosomes; therefore, DNA methylation is their primary means of epigenetic gene regulation.

Three different forms of DNA methylation exist in bacterial genomes: N6-methyladenine (6mA), which is the most prevalent form; N4-methylcytosine (4mC) and 5-methylcytosine (5mC). Although 5mC is the dominant form in eukaryotes, 6mA is the most prevalent form in prokaryotes. DNA is methylated by methyltransferase (MTase) enzymes, which transfer a methyl group from S-adenosyl-L-methionine (SAM) to the appropriate position on target bases (FIG. 1). Importantly, only a select few sequence motifs in each bacterial genome are targeted by MTases; for example, in *Escherichia coli*, 5'-GATC-3' is targeted by DNA adenine methylase (Dam) and 5'-CCWGG-3' by DNA-cytosine methyltransferase (Dcm). However, nearly every occurrence of the target motifs is methylated². The MTase specificity domain that determines the target motif varies widely across species, resulting in a large diversity of methylated motifs across the bacterial kingdom.

MTases function either alongside a cognate restriction enzyme as part of a restriction–modification (RM) system or as 'orphans' that lack a cognate restriction enzyme. DNA methylation mediated by both types of

MTases has been found to play important roles in bacteria^{3–10}. RM systems pre-emptively invading DNA by methylating endogenous cleaving non-methylated foreign DNA. MTases are divided into three main categories on the basis of the subunits involved and the presence of a restriction^{11–14} (FIG. 2). Orphan MTases, Gammaproteobacteria, are thought to regulate replication and gene expression, and mutations⁵. There is also emerging evidence for a role in methylation patterns within bacteria (often caused by phase variation of MTase) and adaptive selection by generating heterogeneity in expression and cellular phenotypes by genetic variation alone^{15,16}. The frequency of the > 6,000 bacterial genomes sequenced to date have been found to encode MTases and are, therefore, likely to be subject to DNA methylation^{16,20}. Nonetheless, the precise sequence targets and biological roles of most MTases remain unknown²¹, largely owing to the historical lack of high-throughput tools for detecting 6mA and 4mC. Indeed, while method development for detecting eukaryotic 5mC has flourished over the past few decades^{22–25}, only modest methodological advances for detecting the principle forms of DNA methylation in bacterial genomes have been made over the same period.

The recent introduction of new sequencing technologies is beginning to address this problem (TABLE 1). In particular, single-molecule, real-time (SMRT) sequencing has enabled all three major forms of bacterial DNA methylation to be detected simultaneously for the first time, and this technology has been used

In order to detect modified nucleotides, IPD values from native DNA sequencing data can be compared with control IPD values from either methylation-free whole-genome amplified (WGA) DNA or precomputed in silico IPD models. The in silico model is

EX2038 (Beaulaurier) p. 159.

Department of Genetics and Genome Sciences and Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
*email: gang.fang@mountsinai.edu
<https://doi.org/10.1038/s41576-018-0087-5>

NATURE REVIEWS | GENETICS

VOLUME 20 | MARCH 2019 | 157

Oxford Nanopore Sequencing Does Not Measure Polymerase Kinetics

Oxford Nanopore Sequencing Does Not Measure Polymerase Kinetics

147. These different sequencing approaches result in different features. For example, the '338 patent discloses interpulse duration (IPD), pulse width (PW), pulse height (PH), local signal-to-noise ratio, dye channel, the χ^2 metric for the measured spectrum, and pulse amplitude variability as possible features (EX1007, 23:14-17, 63:19-22, 66:58-60) whereas Liu compared a 7-feature description of an event to a 57-feature description of an event which have no apparent overlap to the '338 patent features. EX1022 at 3-4. The approaches are not comparable.

148. Moreover, Oxford Nanopore sequencing does not involve polymerase kinetics at all. Dr. Dessimoz's opinion relies on the relationship between sequence context and polymerase kinetics to assert that a POSA would combine nucleotide identity with the polymerase kinetic properties IPD and PW. But, there is no similar or equivalent relationship applicable to Oxford Nanopore sequencing. That nucleotide identity was used in Oxford Nanopore sequencing does not teach or motivate a POSA to use nucleotide identity in PacBio SMRT sequencing or combine nucleotide identity with polymerase kinetic parameters like PW and IPD.

EX2041 (Koo Decl.) ¶¶147-148, pp. 68-69.

Liu Explains that Oxford Nanopore Sequencing Does Not Measure Polymerase Kinetics

bioRxiv preprint doi: <https://doi.org/10.1101/2024.01.21.571616>; this version posted January 21, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Accurate detection of m6A RNA modifications in native RNA sequences

Huanle Liu^{1,2}, Oguzhan Begik^{1,2,3}, Morghan C Lucas^{1,4}, Christopher E. Mason^{5,6,7}, Schraga Schwartz⁸, John S. Mattick^{2,4,6}, Martin A. Smith^{5,9} and Eva Maria Novoa^{1,2,5,4*}

¹Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain

²Department of Neuroscience, Garvan Institute of Medical Research, Darlinghurst, NSW, 2010, Australia

³St-Vincent's Clinical School, UNSW Sydney, Darlinghurst, NSW, 2010, Australia

⁴Universitat Pompeu Fabra (UPF), Barcelona, Spain

⁵Department of Physiology and Biophysics, Weill Cornell Medicine, New York 10021, NY, USA

⁶The Feil Family Brain and Mind Institute, Weill Cornell Medicine, New York 10021, NY, USA

⁷The Human Genome Initiative for Quantitative Prediction, Weill Cornell Medicine, New York 10021, NY, USA

⁸Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

⁹Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst 2010, NSW, Australia

* Present address: Green templeton College, Oxford OX2 6JG, United Kingdom

* Correspondence to: Eva Maria Novoa (eva.novoa@crg.eu)

ABSTRACT

The field of epitranscriptomics has undergone an enormous expansion in the last few years; however, a major limitation is the lack of generic methods to map RNA modifications transcriptome-wide. Here we show that using Oxford Nanopore Technologies, N6-methyladenosine (m6A) RNA modifications can be detected with high accuracy, in the form of systematic errors and decreased base-calling qualities. Our results open new avenues to investigate the universe of RNA modifications with single nucleotide resolution, in individual RNA molecules.

KEYWORDS

RNA modifications, Epitranscriptomics, N6-methyladenosine, Oxford Nanopore Technologies, Direct RNA sequencing

To overcome these limitations, third-generation sequencing technologies, such as the platforms provided by Oxford Nanopore Technologies (ONT) [18] and Pacific Biosciences (PacBio) [13], have been proposed as a new means to detect RNA modifications in native RNA sequences. These technologies can detect RNA modifications by measuring the kinetics of the reverse transcriptase as it encounters a modified RNA -in the case of PacBio-, or by directly sequencing the RNA in its native form by pulling a native RNA through the nanopore -in the case of ONT-. Although ONT direct RNA sequencing is already a reality [19,20], extracting RNA modification information from ONT reads is an unsolved challenge. RNA modifications are known to cause disruptions in the pore current that can be detected upon comparison of raw current intensities – also known as ‘squiggles’ – [18,19]. However, current efforts have not yet yielded an efficient and accurate RNA modification detection algorithm, largely due to the challenges in the alignment and re-squiggle of RNA current intensities.

EX1038 (Liu) p. 2.

Objective Indicia of Non-Obviousness

Nexus Can be Established Two Ways

- 1) [A] patentee is entitled to a rebuttable presumption of nexus between the asserted evidence of secondary considerations and a patent claim if the patentee shows that the asserted evidence is tied to a specific product and the product “is the invention disclosed and claimed.” That is, presuming nexus is appropriate “when the patentee shows that the asserted objective evidence is tied to a specific product and that product ‘embodies the claimed features, and is coextensive with them.’”
- 2) A finding that a presumption of nexus is inappropriate does not end the inquiry into secondary considerations. To the contrary, the patent owner is still afforded an opportunity to prove nexus by showing that the evidence of secondary considerations is the “direct result of the unique characteristics of the claimed invention.”

Fox Factory, Inc. v. SRAM LLC, 944 F.3d 1366, 1373-4 (Fed. Cir. 2019).

Dr. Dessimoz Testified that Tse's Holistic Kinetic Model is Claim 1

4 Q You're welcome.

5 Do you recognize Exhibit 30?

6 A Yes, I believe so.

7 Q What is it?

8 A That's the paper by Tse and Lo that is --

9 essentially corresponds to the '794 patent.

20 Holistic kinetic model, if you go to
21 page 3 of this paper. It's also in the abstract.

22 A Now that you mention it, I recall this is
23 indeed how they refer to their model.

8 Q If you go to the '794 patent, claim 1 --
9 right?

10 A Yes. Yes.

11 Q -- the model is the invention, right?

12 MS. GERRARD: Objection. Form.

13 THE WITNESS: I believe so. Yes.

14 BY MR. WISE:

15 Q Would you agree with me the model is the
16 invention? And that's what claim 1 is about, too,
17 right?

18 MS. GERRARD: Objection. Form.

19 THE WITNESS: Yeah, I believe that the
20 patent -- '794 patent, claim 1, is -- corresponds to
21 the model, this clause in that paper.

Dr. Koo Agrees that Tse's Holistic Kinetic Model is Claim 1

77. I have reviewed 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), which I understand is authored by the inventors of the '794 patent and other collaborators.

In my opinion, Tse's holistic kinetic model is the model of claim 1 of the '794 patent and as shown in EX2054.

EX2041 (Koo Decl.) ¶177, pp. 35-36.

Dr. Koo Showed that the Secondary Considerations Were the Direct Result of the Unique Characteristics of Claim 1

0.55, which is essentially random guessing. The reason the method of claim 1 has such a high AUC is because the inventors discovered, by systematically probing the predictive power of features, an optimal feature set to construct an input data structure that could accurately detect base modification. Those features are a window, and the properties of each nucleotide in the window, namely, the nucleotide identity, the position of the nucleotide with respect to a target position within the window, the PW, and the IPD.

EX2041 (Koo Decl.) ¶39, pp. 15-16.

Objective Indicia: Unexpected Results

Objective Evidence Need Not Mirror the Claim

Objective evidence of nonobviousness need only be “reasonably commensurate with the scope of the claims,” and we do not require a patentee to produce objective evidence of nonobviousness for every potential embodiment of the claim.

Rambus, Inc. v. Rea, 731 F.3d 1248, 1257 (Fed. Cir. 2013)

It is sufficient that the claim recites an element, which is not in the closest prior art, that causes the unexpected result.

Ex Parte Tierney et al., 2013 WL 5399868 at 3

The Expectation of a POSA Was the Method of Claim 1 Would Work as Well as Flusberg, but Experimentation Was Necessary

Perkins
Coie

23 '339 patent.
24 BY MR. WISE:
25 Q My question is, with the Flusberg
1 publications in hand and the Flusberg methodologies
2 in hand, what would a person of skill in the art
3 think would be the result if, instead of fixing the
4 sequence context, as you referred to it, we used
5 instead the input data structure of Lo? What would
6 they have thought the result would be in 2019?
7 MS. GERRARD: Objection. Form.
8 THE WITNESS: I think they would have had
9 a reasonable expectation that Lo's methods would
10 call -- successfully call modification of the bases
11 because it follows very much the teaching that was
12 known in the -- in the prior art records, including
13 the importance of the context, the interpulse
14 duration, pulse width, and that entity of the bases.

15 And, therefore, I think a person of
16 ordinary skill in the art in 2019 would have
17 expected, you know, Lo's approach to perform at
18 least as well as the publication that was -- that
19 came out nine years prior, using, you know, data
20 that was necessarily much more rudimentary than what
21 happened during this period. That would be the
22 expectation.
23 But, again, you know, the expectation is
24 one thing. But usually a person of skill in the art
25 runs the experiment, and this is what then provides
1 a kind of more definitive indication.

EX2037 (Dessimoz Dep.) 112:23-114:1.

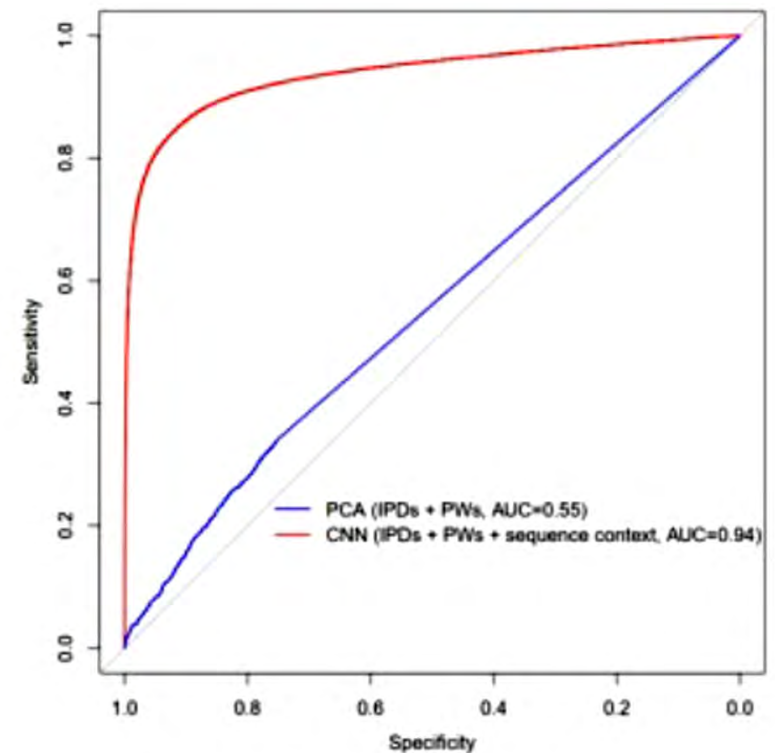
The Claim 1 Method is Unexpectedly Better Than Flusberg

164. The surprising and unexpected result of the method of claim 1 is its very high prediction (0.94) compared to the Flusberg method (0.55) which is essentially random guessing.

EX2041 (Koo Decl.) ¶ 164, p. 77.

FIG. 29 shows that the AUC of the method based on the principal component analysis for which two principal components were used in Flusberg et al.'s study (Flusberg et al., 2010) involving IPDs and PWs were much less accurate (AUC: 0.55) than the approach based on a convolutional neural network involving IPD and PW as well as sequence context as shown in our disclosure (AUC: 0.94).

EX1001 ('794 Patent) FIG. 29, 35:38-44.



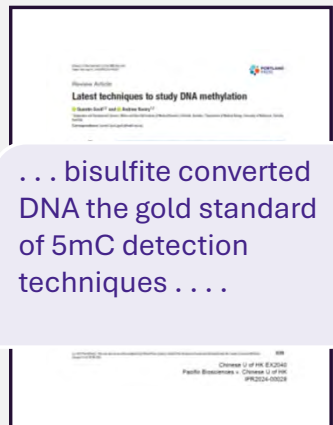
Objective Indicia: Long Felt Need/Failure of Others

Bisulfite Sequencing Is the Recognized Gold Standard DNA Methylation Detection Method



... the gold standard for characterizing 5mC ... , bisulfite sequencing ...

– **EX2038 at 158**
Beaulaurier (2019)



... bisulfite converted DNA the gold standard of 5mC detection techniques ...

– **EX2040 at 640**
Gouil (2019)



... bisulfite sequencing is widely used ... as a gold standard ...

– **EX2042 at 441**
Sang (2021)



Whole genome bisulfite sequencing ... is currently the gold standard ...

– **EX2039 at 15**
Birknerova (2022)



BS sequencing ... [is] gold standard[]. ...

– **EX2043 at 10**
Wang (2023)

metry^{16,17,23} and nanopore amperometry²⁴. To our knowledge, however, no high-throughput method has been demonstrated that allows the determination of primary sequence at the same time as methylation status. Here we present a method to directly detect DNA methylation during single-molecule, real-time (SMRT) DNA sequencing, a technique for studying nucleic acid sequence and structure²⁵. In this technique, single DNA polymerase molecules are observed in real time while they catalyze the incorporation of fluorescently labeled nucleotides complementary to a template nucleic acid strand. These reactions are measured

PACIFIC BIOSCIENCES OF CALIFORNIA, INC. EXHIBIT - 1007

Schadt Recognized that the Unsolved Need for Direct Detection of DNA Methylation Through SMRT Sequencing Without Bisulfite Conversion

Method

Modeling kinetic rate variation in third generation

Presently, only 5-mC can be routinely detected using NGS technology and then only after samples have been bisulfite treated, a procedure that creates a number of experimental and informatics challenges (Cokus et al. 2008). Thus, new technologies and assays are needed to fully characterize DNA and RNA sequences if we hope to completely characterize variation in genomes and transcriptomes.

every known type, providing a path forward for detecting base modifications as a routine part of sequencing. However, to date no statistical framework has been proposed to enhance the power to detect these events while also controlling for false-positive events. By modeling enzyme kinetics in the neighborhood of an arbitrary location in a genomic region of interest as a conditional random field, we provide a statistical framework for incorporating kinetic information at a test position of interest as well as at neighboring sites that help enhance the power to detect KV events. The

Presently, 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. In the first study published demonstrating the utility of kinetic variation information for detecting chemical modifications like 5-mC, a simple ratio measure was used to compare the kinetic parameters between two experimental groups for a given base position of interest (Flusberg et al. 2010). Empirically determined

PACIFIC BIOSCIENCES OF CALIFORNIA, INC. EXHIBIT - 1010

John Beaulaurier, Eric E. Schadt^{ID} and Gang Fang^{ID}*

using modern sequencing technologies

John Beaulaurier, Eric E. Schadt^{ID} and Gang Fang^{ID}*

Abstract | Prokaryotic DNA contains three forms of modification: N6-methyladenine (m6A),

However, detection of 5mC and 5hmC requires either high sequencing depth or additional steps to convert them to 5-formylcytosine and 5caC, both of which have higher signal-to-noise ratios^{48,51}.

diversity is a closely regulated process and genes for many of the enzymes involved in DNA methylation are highly conserved.

Adaptive selection
An evolutionary process through which natural selection favors genetic changes that lead to a fitness advantage over other genotypes.

Department of Genetics and Genome Sciences and Center for Genome Sciences and Policy, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

*email: gang.fang@unc.edu
genetics.unc.edu
DOI:10.1016/j.celrep.2018.03.020

NATURE REVIEWS GENETICS

REVIEWS

VOLUME 19 | MARCH 2018
Chinese U of HK EX2038
Pacific Biosciences v. Chinese U of HK
IPR2024-00028

By the Time of the Invention, No One Had Solved the Need for Direct Detection of DNA Methylation through SMRT Sequencing Without Bisulfite Conversion

Perkins
Coie

Essays in Biochemistry (2019) 63:639–649
https://doi.org/10.1042/ESB20190027



Review Article

Latest techniques to study DNA methylation

Quentin Gouil^{1,2} and Andrew Keniry^{1,2}

¹Genomics and Development Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Australia; ²Department of Medical Biology, University of Melbourne, Parkville, Australia

Correspondence: Quentin Gouil (gouil@wehi.edu.au)



Bisulfite sequencing is a powerful technique to detect 5-methylcytosine in DNA that has immensely contributed to our understanding of epigenetic regulation in plants and animals. Meanwhile, research on other base modifications, including 6-methyladenine and 4-methylcytosine that are frequent in prokaryotes, has been impeded by the lack of a comparable technique. Bisulfite sequencing also suffers from a number of drawbacks that are difficult to surmount, among which DNA degradation, lack of specificity, or short reads with low sequence diversity. In this review, we explore the recent refinements to bisulfite sequencing protocols that enable targeting genomic regions of interest, detecting derivatives of 5-methylcytosine, and mapping single-cell methylomes. We then present the unique advantage of long-read sequencing in detecting base modifications in native DNA and highlight the respective strengths and weaknesses of PacBio and Nanopore sequencing for this application. Although analysing epigenetic data from long-read platforms remains challenging, the ability to detect various modified bases from a universal sample preparation, in addition to the mapping and phasing advantages of the longer read lengths, provide long-read sequencing with a decisive edge over short-read bisulfite sequencing for an expanding number of applications across kingdoms.

Because the signal-to-noise ratio is low, the detection of base modifications in single-molecule is inaccurate and often requires summarising at the genomic position-level. As 6mA and 4mC produce strong kinetic signatures, a coverage of $25 \times$ per strand is recommended [68]. However the subtle effects of 5mC and 5hmC increase the requirements to $250 \times$, unless they are enriched or modified to produce a larger kinetic effect by glycosylation or TET-conversion to 5-carboxylcytosine [68,69]. PacBio sequencing therefore only achieves single-molecule resolution for certain marks and on relatively short fragments (≤ 2 kb) that can be read a large number of times by the polymerase [70]. With longer fragments, cell-to-cell variability cannot be investigated in detail. PacBio's price per Gb, sensitivity to particular modifications, resolution and high coverage requirements make this technology particularly suited to bacterial genomes, where 6mA and 4mC are frequent and often concentrated on specific motifs (Figure 1). The use of SMRT

IPR2024-00028

Published online 22 February 2019

Nucleic Acids Research, 2019, Vol. 47, No. 8 e46
doi:10.1093/nar/gkz107

Using long-read sequencing to detect imprinted DNA

(bioRxiv doi:10.1101/385849). We note that, for the dominant eukaryotic genome base modification at 5mC, the PacBio technology requires very high coverage making it impractical for use in the analysis of mammalian genomes (23). PacBio SMRT sequencing can be combined with bisulfite treatment (SMRT-BS) to facilitate 5mC detection, but this approach is currently only available for targeted sequencing (24) and the bisulfite treatment introduces the same drawbacks noted above in addition to fragmenting the DNA. Additionally, while PacBio technology is limited

lotype from which each read arises. The long-read property is exploited to characterize, using novel methods, both methylation and haplotype for reads that have reduced basecalling precision compared to Sanger sequencing. We validate the analysis both through comparison of nanopore-derived methylation patterns with those from Reduced Representation Bisulfite Sequencing data and through comparison with previously reported data. Our analysis successfully identifies known imprinting control regions (ICRs) as well as some novel differentially methylated regions which, due to their proximity to hitherto unknown monoallelically expressed genes, may represent new ICRs.

genes (6). In accordance with its primordial role in allocation of resources from the mother to the offspring, the placenta, along with the brain, is the organ where parental conflict results in the most pronounced imprinted expression (7–9). We thus conduct a survey of differential methylation and expression in murine embryonic placenta.

Recent studies have increased the number of genes identified as subject to imprinting in mouse to ~200 (10–15). The cause of the differential expression between paternal and maternal alleles is only known for a subset of these genes; maternal histone marks can play a role (14), and in other cases it involves the differential methylation of adjacent regions (5). The differential methylation patterns may be established in the gametes and persist through the epigenetic reprogramming occurring after fertilization (16).

To whom correspondence should be addressed. Tel: +61 3 9345 2856; Fax: +61 9347 0852; Email: methinfo@wehi.edu.au

Correspondence may also be addressed to Scott Gigante; Email: scott.gigante@perkinscoie.com

¹The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

²The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint last authors.

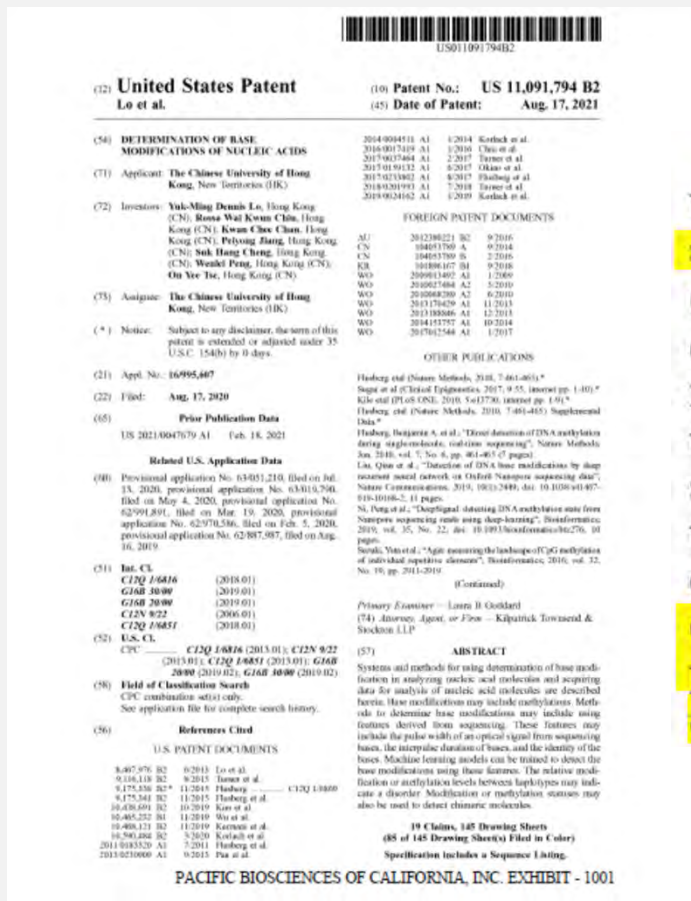
© The Author(s) 2019. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Chinese U of HK EX2051

Pacific Biosciences v. Chinese U of HK

IPR2024-00028

The '794 Patent Explains the Long-felt Need for Direct Detection of DNA Methylation Through SMRT Sequencing Without Bisulfite Conversion



One disadvantage of such bisulfite-based approaches is that the bisulfite conversion step has been reported to significantly degrade the majority of the treated DNA (Grunau, 2001). Another disadvantage is that the bisulfite conversion step would create strong CG biases (Olova et al., 2018), resulting in the reduction of signal-to-noise ratios typically for DNA mixtures with heterogeneous methylation states. Furthermore, bisulfite sequencing would not be able to sequence long DNA molecules because of the degradation of DNA during bisulfite treatment. Thus, there is a need to determine the modification of bases of nucleic acids, without prior chemical (e.g. bisulfite conversion) and nucleic acid amplification (e.g. using the PCR).

EX1001 ('794 Patent) 2:19-31.

The Solution to the Long-felt Need for Direct Detection of DNA Methylation Through SMRT Sequencing Without Bisulfite Conversion Was a Direct Result of the Unique Characteristics of Claim 1

8 Q Did you look at the data to see how the
9 data on the methodology of Lo, using the input data
10 structure that is the focus of this dispute, and how
11 well it matched up to bisulfite sequencing?

12 MS. GERRARD: Objection. Form.

13 THE WITNESS: I do not recall specifically
14 looking at this -- this side-by-side comparison.

15 BY MR. WISE:

16 Q If you go to column 48, line 22, it says
17 (as read), "As shown in Figure 51, there was a good
18 correlation ($r:0.96$, $p\text{-value } 0.0023$) between overall
19 methylation levels deduced by single molecule,
20 real-time sequencing with the present methylation
21 prediction approach and bisulfite sequencing."
22 Would you call that scientifically
23 successful?

24 MS. GERRARD: Objection. Form.

25 THE WITNESS: Well, so we need to look at
1 the particular data and how difficult this was.
2 But on the face of this good correlation,
3 I believe this could be -- this could be qualified
4 as a success, yes.

The Method of Claim 1 Solved the Long-felt Need for Direct Detection of DNA Methylation Through SMRT Sequencing Without Bisulfite Conversion

As explained above, the '794 patent tested various features as properties to be used in the input data structure and concluded that the combination of IPD, PW, and nucleotide identity and relative position to a target position in a window was the most predictive feature set of DNA methylation detection. Using these features, the identity of the nucleotide, the position of the nucleotide relative to a target position in a window, the PW corresponding to the nucleotide and the IPD, the '794 patent inventors, the Lo inventors, trained their model according to the method of claim 1. EX2002 at ¶¶44-48, EX1001 at 30:60-31:63. Using the trained model with the input data structure of claim 1, as explained above, the Lo inventors conducted a series of experiments that compared their method to the “gold standard” –bisulfite sequencing and showed good correlation ($r = 0.90$ or higher) for most of the tested biological samples. (See ¶¶59-72 above). Dr. Lo's team solved the long-felt need for direct detection of DNA methylation, without bisulfite conversion, through SMRT sequencing.

EX2041 (Koo Decl.) ¶158, pp. 74-75.

Objective Indicia: Industry Praise

PacBio CEO Christian Henry Praised the Invention

First, we'd like to congratulate Dr. Dennis Lo and colleagues at the Chinese University of Hong Kong 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. 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. We'd also like to congratulate the team at HudsonAlpha Institute of Biology

EX2017 (Dec. 31, 2020 Earnings Call) p. 5.

PacBio President of Segment Marketing Jennifer Stone Praised the Invention

Perkins
Coie

Speaker: First, I'd like to highlight the great work coming out of Dennis Lo's lab at The Chinese University of Hong Kong earlier this year. The team developed a methodology to directly examine 5-methy-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, Professor Lo et al. 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.

EX2020 & 2023 (2021 ASHG annual meeting video and transcript).

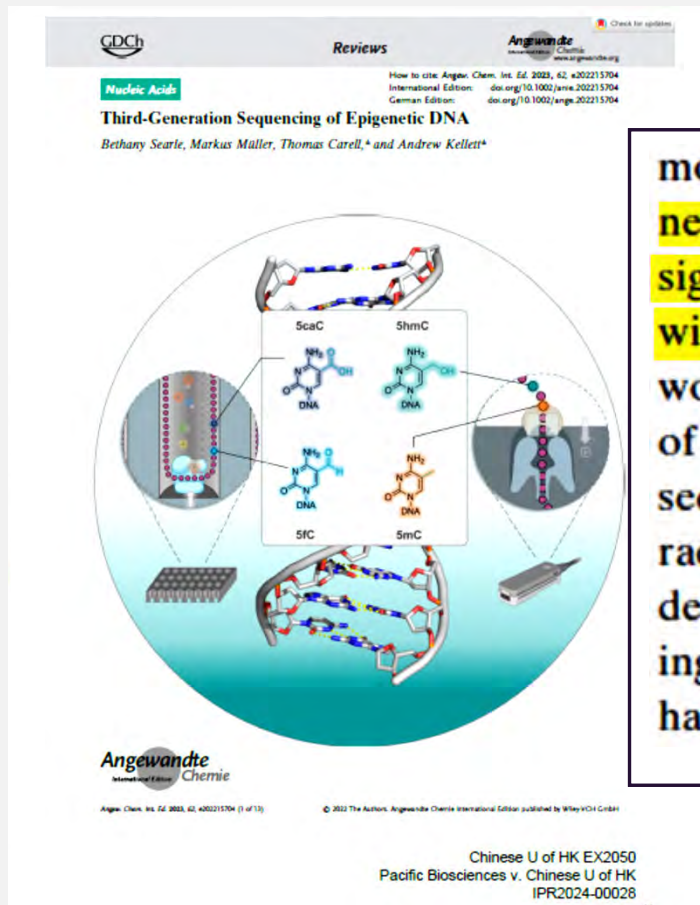
The Industry Praised the Invention



Pulse signals in SMRT sequencing, which are associated with nucleotide polymerization, include the interpulse duration (IPD) and the pulse width (PW). IPD represents the time interval between two consecutive base incorporations and PW is characterized by the emission signal related to base incorporation [88,91]. Because the changes in the kinetic signal caused by 5mC modification are extremely subtle, robust genome-wide measurement of 5mC modification is very challenging [91]. The first study concerning detection of 5mC using SMRT sequencing took advantage of the enhancement of the kinetic signature of 5-carboxylcytosine upon conversion of 5mCs using Tet1 [92]. Following this, Tse et al. [91] attempted to develop a holistic kinetic (HK) model to improve the accurate detection of 5mC using SMRT sequencing. Based on the validation datasets generated using amplified DNA and DNA treated with M.SssI (the CpG methyltransferase which methylates the C5 position of all CpG sites in a double-stranded DNA), the HK model dramatically improved 5mC detection rates by 90% at 94% specificity. This model was used for different types of extracted DNA, such as buffy coats, placental tissues, and tumor tissues. The overall methylation level analyzed by the HK model highly correlated with bisulfite sequencing results (99%); therefore, the HK model constitutes a new, viable approach for studying epigenetic modifications in molecular diagnostic applications.

EX2039 (Birknerova) p. 16.

The Industry Praised the Invention



modifications has proven challenging. Recently, a neural network was used to identify 5mC in native DNA, which significantly improved detection generating 99 % agreement with bisulfite sequencing data.^[61] Uniquely, the neural network is able to simultaneously account for multiple variables of polymerase kinetics, including IPDs, pulse width and sequence context simultaneously, leading to greater accuracy. At present, this does not extend beyond the direct detection of 5mC and 5hmC, analogous to bisulfite sequencing. The detection of further oxidation of cytosine residues has yet to be achieved.

EX2050 (Searle) p. 7.

Objective Indicia: Copying

PacBio Had Access to the Tse Article

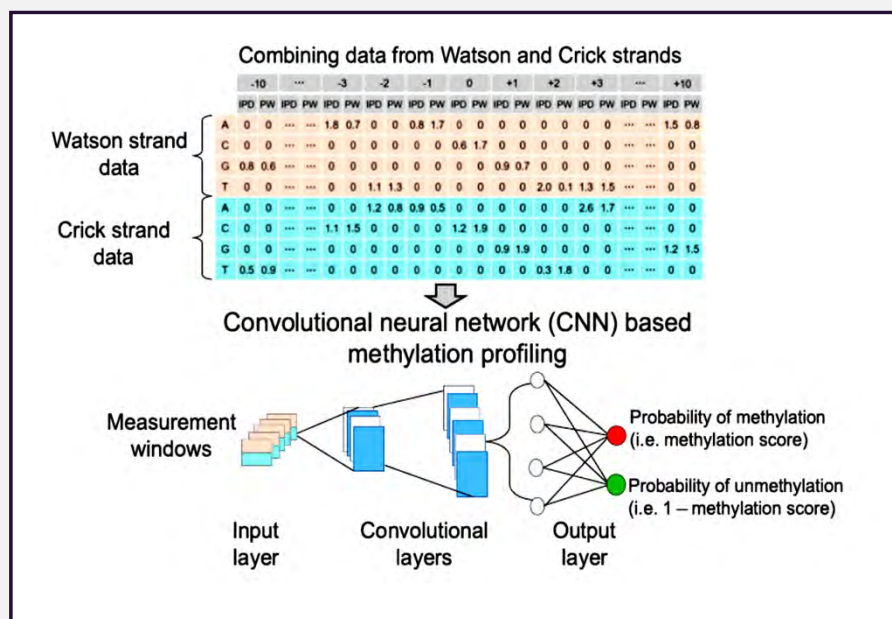
First, we'd like to congratulate Dr. Dennis Lo and colleagues at the Chinese University of Hong Kong on the development of a new method for highly accurate simultaneous determination of DNA sequence and CPG methylation in one go, published in the PNAS Journal in January. This new method will help researchers 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. We'd also like to congratulate the team at HudsonAlpha Institute of Biology

EX2017 (Dec. 31, 2020 Earnings Call) p. 5.

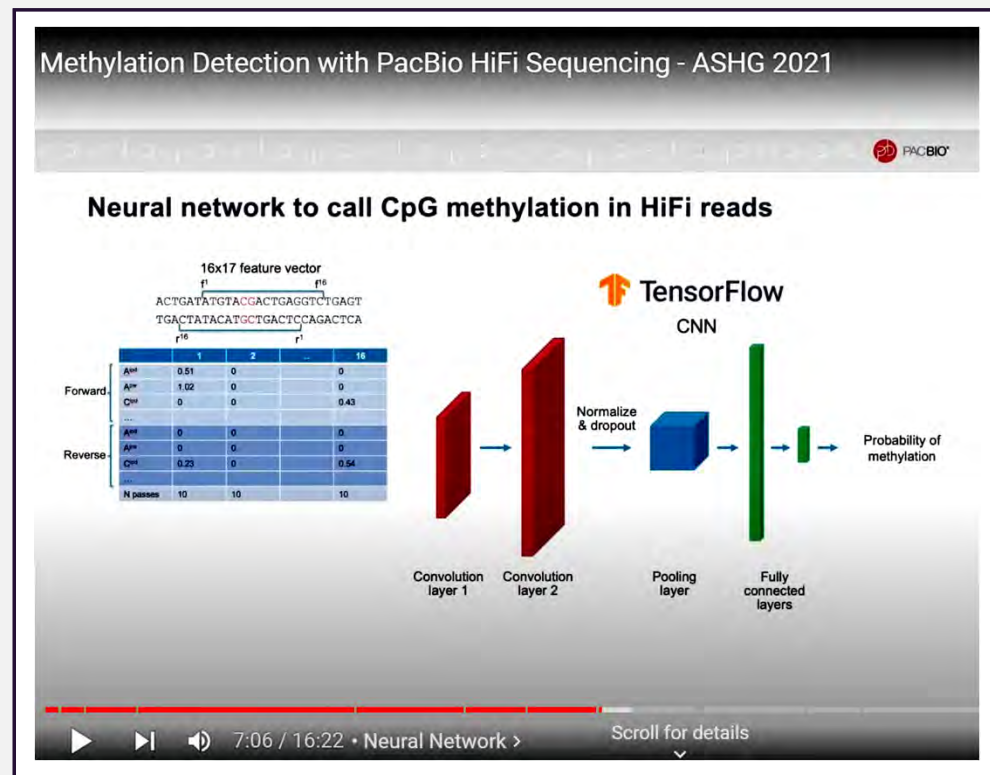
The PacBio Product Bears a Remarkable Similarity to Holistic Kinetic Model Disclosed in the Tse Article, Fig. 1

Perkins
Coie

EX1030 (Tse) Fig. 1 (partial).



EX2024 (Wenger) screenshot at 7:06.



EX2041 (Koo Decl.) ¶¶ 80-81.

PacBio Copied Tse



Warren A. Cheung¹, Adam F. Johnson¹, William J. Rowell², Emily Farrow^{1,3}, Richard Hall², Ana S. A. Cohen^{3,4}, John C. Means¹, Tricia N. Zion¹, Daniel M. Portik², Christopher T. Saunders², Boryana Koseva¹, Chengpeng Bi¹, Tina K. Truong⁵, Carl Schwendinger-Schreck¹, Byunggil Yoo¹, Jeffrey J. Johnston¹, Margaret Gibson¹, Gilad Evrony⁵, William B. Rizzo⁶, Isabelle Thiffault^{3,4}, Scott T. Younger^{1,3}, Tom Curran⁷, Aaron M. Wenger², Elin Grundberg^{1,3} & Tomi Pastinen^{1,3}

The technical ability to interrogate mCpG status from single-molecular real-time (SMRT) sequencing (kinetic) data was recently demonstrated using the HiFi-GS long-read platforms²⁶. This ability has now been integrated in the HiFi-GS production pipeline (Sequel IIe system release v11, PacBio, Menlo Park, CA) that combines a deep-learning model integrating sequencing kinetics and base context, to generate mCpG profiles genome-wide from standard sequencing libraries. The augmented 5-base HiFi-GS platform allows single-

EX2044 (Cheung) p. 2.