

BRIEF SUMMARY

[0007] We have developed a new method that, in one ~~embodiment~~example, allows the determination of base modifications, such as 5mC in nucleic acids without template DNA pre-treatment such as enzymatic and/or chemical conversions, or protein and/or antibody binding.

5 While such template DNA pre-treatment is not necessary for the determination of the base modifications, in examples that are shown, certain pre-treatment (e.g. digestion with restriction enzymes) may serve to enhance ~~aspects~~examples of the ~~invention~~disclosure (e.g. allowing the enrichment of CpG sites for analysis). The ~~embodiments~~examples present in this disclosure could be used for detecting different types of base modification, for example, including but not
10 limited to 4mC, 5hmC, 5fC, and 5caC, 1mA, 3mA, 7mA, 3mC, 2mG, 6mG, 7mG, 3mT, and 4mT, etc. Such ~~embodiment~~examples can make use of features derived from sequencing, such as kinetic features, that are affected by the various base modifications, as well as an identity of nucleotides in a window around a target position whose methylation status is determined.

[0008] ~~Embodiments~~Examples of the present ~~invention~~disclosure can be used for, but is not
15 limited to, single molecule sequencing. One type of single molecule sequencing is single molecule, real-time sequencing in which the progress of the sequencing of a single DNA molecule is monitored in real-time. One type of single molecule, real-time sequencing is that commercialized by Pacific Biosciences using their Single Molecule, Real-Time (SMRT) system. Methods may use the pulse width of a signal from sequencing bases, the interpulse duration
20 (IPD) of bases, and the identity of the bases in order to detect a modification in a base or in a neighboring base. Another single molecule system is that based on nanopore sequencing. One example of a nanopore sequencing system is that commercialized by Oxford Nanopore Technologies.

[0009] The methods we have developed can serve as tools to detect base modifications in
25 biological samples to assess the methylation profiles in the samples for various purposes including but not limited to research and diagnostic purposes. The detected methylation profiles can be used for different analysis. The methylation profiles can be used to detect the origin of DNA (e.g., maternal or fetal, tissue, bacterial, or DNA obtained from tumor cells enriched from the blood of a cancer patient). Detection of aberrant methylation profiles in tissues aids the

identification of developmental disorders in individuals, identify and prognosticate tumors or malignancies.

[0010] ~~Embodiment~~Examples of the present inventiondisclosure may include analyzing the relative methylation levels of haplotypes of an organism. An imbalance in the methylation levels between the two haplotypes may be used to determine a classification of a disorder. A higher imbalance may indicate the presence of a disorder or a more severe disorder. The disorder may include cancer.

[0011] Methylation patterns in a single molecule can identify chimera and hybrid DNA. Chimeric and hybrid molecules may include sequences from two different genes, chromosomes, organelles (e.g. mitochondria, nucleus, chloroplasts), organisms (mammals, bacteria, viruses, etc.), and/or species. Detecting junctions of chimeric or hybrid DNA molecules may allow for detecting gene fusions for various disorders or diseases, including cancer, prenatal, or congenital disorders.

[0012] A better understanding of the nature and advantages of ~~embodiment~~examples of the present inventiondisclosure may be gained with reference to the following detailed description and the accompanying drawings.

[0012a] In one aspect, disclosed herein is a method for detection of a methylation of a nucleotide in a nucleic acid molecule, the method comprising: receiving a first plurality of first data structures, each first data structure of the first plurality of first 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 the respective window comprises at least two nucleotides, wherein each of the first nucleic acid molecules is sequenced by measuring pulses in an optical signal corresponding to the nucleotides, wherein the methylation has a known first state in the nucleotide at a target position in each window of each first nucleic acid molecule, each first 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 the 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, 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 for the methylation of the nucleotide at the target position; and training a model using the plurality of first training samples by optimizing 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 methylation.

[0012b] In another aspect, disclosed herein is a computer product comprising a non-transitory computer readable medium storing a plurality of instructions that when executed control a computer system to perform the method of the disclosure.

[0012c] In another aspect, disclosed herein is a system comprising: the computer product of the disclosure; and one or more processors for executing instructions stored on the computer readable medium.

[0012d] In another aspect, disclosed herein is a method for detecting a methylation of a nucleotide in a nucleic acid molecule, the method comprising: receiving input data acquired by measuring pulses in an optical signal corresponding to a window of nucleotides in a sample nucleic acid molecule, and wherein the window comprises at least two nucleotides, the input data 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; determining whether the methylation is present in a nucleotide at the target position within the window using: for each nucleotide within the window: the identity of the nucleotide, the position of the nucleotide with respect to the target position, the width of the pulse corresponding to the nucleotide, and the interpulse duration.

[0012e] In another aspect, disclosed herein is a computer product comprising a non-transitory computer readable medium storing a plurality of instructions that when executed control a computer system to perform the method of the disclosure.

[0012f] In another aspect, disclosed herein is a system comprising: the computer product of the disclosure; and one or more processors for executing instructions stored on the computer readable medium.

5

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 illustrates SMRT sequencing of molecules carrying base modifications according to embodiments of the present invention.

[0014] FIG. 2 illustrates SMRT sequencing of molecules carrying methylated and unmethylated CpG sites according to embodiments of the present invention.

10

[0015] FIG. 3 illustrates interpulse durations and pulse width according to embodiments of the present invention.

[0016] FIG. 4 shows an example of a measurement window of the Watson strand of DNA for detecting a base modification according to embodiments of the present invention.

WHAT IS CLAIMED IS:

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

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;~~

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

for each nucleotide within the window:

an identity of the nucleotide,

a position of the nucleotide with respect to the 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,
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 for the
methylation 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;
determining, using the model, whether the modification is present in a nucleotide
at the target position within the window in the input data structure
training a model using the plurality of first training samples by optimizing
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 methylation.

2. — The method of claim 1, wherein:

the input data structure is one input data structure of a plurality of input
data structures;
the sample nucleic acid molecule is one sample nucleic acid molecule of a
plurality of sample nucleic acid molecules;
the plurality of sample nucleic acid molecules are obtained from a
biological sample of a subject, and
each input data structure corresponds to a respective window of
nucleotides sequenced in a respective sample nucleic acid molecule of the plurality of
sample nucleic acid molecules, and
the method further comprising:
receiving the plurality of input data structures,
inputting the plurality of input data structures into the model, and

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

1 3.——The method of claim 2, further comprising:
2 determining the modification is present at one or more nucleotides, and
3 determining a classification of a disorder using the presence of the modification at
4 one or more nucleotides.

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

1 5.——The method of claim 3, further comprising:
2 determining that the classification of the disorder is that the subject has the
3 disorder, and
4 treating the subject for the disorder.

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

1 7.——The method of claim 2, further comprising:
2 determining the modification is present at one or more nucleotides, and
3 determining a clinically relevant DNA fraction, a fetal methylation profile, a
4 maternal methylation profile, a presence of an imprinting gene region, or a tissue of origin using
5 the presence of the modification at one or more nucleotides.

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

1 9.——The method of claim 2, wherein:
2 the plurality of sample nucleic acid molecules align to a plurality of genomic
3 regions,
4 for each genomic region of the plurality of genomic regions:
5 a number of sample nucleic acid molecules is aligned to the genomic
6 region,
7 the number of sample nucleic acid molecules is greater than a cutoff number.

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

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

1 12.—The method of claim 1, wherein:
2 the window of nucleotides corresponding to the input data structure comprises
3 nucleotides on a first strand of the sample nucleic acid molecule and nucleotides on a second
4 strand of the sample nucleic acid molecule, and
5 the input data structure further comprises for each nucleotide within the window a
6 value of a strand property, the strand property indicating the nucleotide being present on either
7 the first strand or the second strand.

1 13.—The method of claim 12, wherein the sample nucleic acid molecule is a
2 circular DNA molecule formed by:
3 cutting a double stranded DNA molecule using a Cas9 complex to form a cut
4 double stranded DNA molecule, and
5 ligating a hairpin adaptor onto an end of the cut double stranded DNA molecule.

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

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

1 16.—The method of claim 15, wherein each nucleotide within the window is
2 enriched by:
3 cutting a double stranded DNA molecule using a Cas9 complex to form a
4 cut double stranded DNA molecule, and ligating a hairpin adaptor onto an end of the cut
5 double stranded DNA molecule, or
6 filtered by:

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

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

1 ~~18. — A method for detection of a modification of a nucleotide in a nucleic acid~~
2 ~~molecule, the method comprising:~~

3 ~~receiving a first plurality of first data structures, each first data structure of the~~
4 ~~first plurality of first data structures corresponding to a respective window of nucleotides~~
5 ~~sequenced in a respective nucleic acid molecule of a plurality of first nucleic acid molecules,~~
6 ~~wherein each of the first nucleic acid molecules is sequenced by measuring pulses in an optical~~
7 ~~signal corresponding to the nucleotides, wherein the modification has a known first state in the~~
8 ~~nucleotide at a target position in each window of each first nucleic acid molecule, each first data~~
9 ~~structure comprising values for the following properties:~~

10 ~~for each nucleotide within the window:~~

11 ~~an identity of the nucleotide,~~

12 ~~a position of the nucleotide with respect to the target position~~
13 ~~within the respective window,~~

14 ~~a width of the pulse corresponding to the nucleotide, and~~

15 ~~an interpulse duration representing a time between the pulse~~
16 ~~corresponding to the nucleotide and a pulse corresponding to a neighboring~~
17 ~~nucleotide,~~

18 ~~storing a plurality of first training samples, each including one of the first plurality~~
19 ~~of first data structures and a first label indicating the first state for the modification of the~~
20 ~~nucleotide at the target position; and~~

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

192. The method of claim 181, further comprising:
receiving a second plurality of second data structures, each second data structure
of the second plurality of second data structures corresponding to a respective window of
nucleotides sequenced in a respective nucleic acid molecule of a plurality of second nucleic acid
molecules, wherein the ~~modification-methylation~~ has a known second state in a nucleotide at a
target position within each window of each second nucleic acid molecule, each second data
structure comprising values for the same properties as the first plurality of first data structures;
storing a plurality of second training samples, each including one of the second
plurality of second data structures and a second label indicating the second state of the nucleotide
at the target position;
wherein ~~training~~:
the first state or the second state is that the ~~modification-methylation~~ is
present and the other state is that the ~~modification-methylation~~ is absent, and
training the model further comprises using the plurality of second training
samples by optimizing parameters of the model based on outputs of the model matching
or not matching corresponding labels of the second labels when the second plurality of
second data structures are input to the model.

203. The method of claim 192, wherein the plurality of first nucleic acid
molecules is the same as the plurality of the second nucleic acid molecules.

214. The method of claim 192, ~~wherein the modification comprises~~
~~methylation~~, wherein the plurality of first nucleic acid molecules is generated using multiple
displacement amplification with methylated nucleotides of a first type, and wherein the plurality
of second nucleic acid molecules is generated using multiple displacement amplification with
unmethylated nucleotides of the first type.

225. The method of claim 181, wherein the optical signal is a fluorescence
signal from a dye-labeled nucleotide.

1 236. The method of claim 181, wherein each window associated with the first
2 plurality of first data structures comprises 4 consecutive nucleotides on a first strand of each first
3 nucleic acid molecule.

1 247. The method of claim 236, wherein the windows associated with the first
2 plurality of first data structures comprise the same number of consecutive nucleotides.

1 258. The method of claim 181, wherein:
2 each window associated with the first plurality of data structures comprises
3 nucleotides on a first strand of the first nucleic acid molecule and nucleotides on a second strand
4 of the first nucleic acid molecule, and
5 each first data structure further comprises for each nucleotide within the window a
6 value of a strand property, the strand property indicating the nucleotide being present on either
7 the first strand or the second strand.

1 269. The method of claim 181, wherein the neighboring nucleotide is the
2 adjacent nucleotide.

1 2710. The method of claim 181, wherein the width of the pulse is the width of
2 the pulse at half the maximum value of the pulse.

1 2811. The method of claim 181, wherein the interpulse duration is the time
2 between the maximum value of the pulse associated with the nucleotide and the maximum value
3 of the pulse associated with the neighboring nucleotide.

1 2912. The method of claim 181, wherein the model comprises a convolutional
2 neural network comprising:

3 a set of convolutional filters configured to filter the first plurality of first data
4 structures,

5 an input layer configured to receive the filtered first plurality of first data
6 structures,

7 a plurality of hidden layers including a plurality of nodes, a first layer of the
8 plurality of hidden layers coupled to the input layer; and

9 an output layer coupled to a last layer of the plurality of hidden layers and
10 configured to output an output data structure, the output data structure comprising the properties.

1 ~~30. The method of claim 18, wherein the modification comprises a~~
2 ~~methylation of the nucleotide at the target position.~~

1 ~~31~~13. The method of claim ~~30~~1, wherein the known first states include a
2 methylated state for a first portion of the first data structures and an unmethylated state for a
3 second portion of the first data structures.

1 ~~32~~14. The method of claim ~~30~~1, wherein the methylation comprises 4mC (N4-
2 methylcytosine), 5mC (5-methylcytosine), 5hmC (5-hydroxymethylcytosine), 5fC (5-
3 formylcytosine), 5caC (5-carboxylcytosine), 1mA (N1-methyladenine), 3mA (N3-
4 methyladenine), 6mA (N6-methyladenine), 7mA (N7-methyladenine), 3mC (N3-
5 methylcytosine), 2mG (N2-methylguanine), 6mG (O6-methylguanine), 7mG (N7-
6 methylguanine), 3mT (N3-methylthymine), or 4mT (O4-methylthymine).

15. The method of claim 1, wherein the methylation comprises 5mC (5-
methylcytosine) or 6mA (N6-methyladenine).

16. The method of claim 1, wherein the methylation comprises 5mC (5-
methylcytosine).

1 ~~33. The method of claim 18, wherein the modification comprises an oxidation~~
2 ~~change.~~

1 ~~34~~17. The method of claim ~~18~~1, wherein each data structure further comprises a
2 value for a height of the pulse corresponding to each nucleotide within the window.

1 ~~35~~18. The method of claim ~~18~~1, wherein the optical signal corresponding to the
2 nucleotides results from the nucleotides or a tag associated with the nucleotides.

1 ~~36~~19. The method of claim 18, wherein each target position is the center of the
2 respective window.

1 3720. The method of claim ~~18~~1, wherein the methylation is absent in each
2 window of each first nucleic acid molecule.

1 3821. The method of claim ~~18~~1, wherein:
2 each first data structure of the first plurality of first data structures excludes first
3 nucleic acid molecules with an interpulse duration or a width of a pulse below a cutoff value.

1 3922. The method of claim ~~18~~1, wherein:
2 ~~the modification comprises a methylation, and~~
3 the plurality of first training samples is generated by:
4 amplifying a plurality of nucleic acid molecules using a set of nucleotides,
5 wherein the set of nucleotides includes 6mA at a specified ratio.

1 4023. The method of claim 3922, wherein the methylation comprises 6mA (N6-
2 methyladenine).

1 4124. The method of claim ~~1 or claim 18~~, wherein at least some of the plurality
2 of first nucleic acid molecules each include a first portion corresponding to a first reference
3 sequence and a second portion corresponding to a second reference sequence that is disjoint from
4 the first reference sequence.

1 4225. The method of claim ~~1 or claim 18~~, further comprising:
2 validating the model using a plurality of chimeric nucleic acid molecules, each
3 including a first portion corresponding to a first reference sequence and a second portion
4 corresponding to a second reference sequence, wherein the first portion has a first methylation
5 pattern and the second portion has a second methylation pattern.

1 4326. The method of claim ~~4124~~ or claim 4225, wherein the first portion is
2 treated with a methylase.

1 4427. The method of claim 4326, wherein the second portion corresponds to an
2 unmethylated portion of the second reference sequence.

1 4528. The method of claim 4124 or claim 4225, wherein the first reference
2 sequence is human, and wherein the second reference sequence is from a different animal.

1 46. ~~A method of analyzing a biological sample of an organism having a first~~
2 ~~haplotype and a second haplotype in a first chromosomal region, the biological sample including~~
3 ~~DNA molecules, the method comprising:~~

4 ~~analyzing a plurality of DNA molecules from the biological sample, wherein~~
5 ~~analyzing a DNA molecule includes:~~

6 ~~identifying a location of the DNA molecule in a reference human genome;~~
7 ~~determining a respective allele of the DNA molecule; and~~
8 ~~determining whether the DNA molecule is methylated at one or more genomic~~
9 ~~sites;~~

10 ~~identifying one or more heterozygous loci of a first portion of the first~~
11 ~~chromosomal region, each heterozygous locus including a corresponding first allele in the first~~
12 ~~haplotype and a corresponding second allele in the second haplotype;~~

13 ~~identifying a first set of the plurality of DNA molecules that each:~~
14 ~~is located at any one of the one or more heterozygous loci,~~
15 ~~includes the corresponding first allele of the heterozygous locus, and~~
16 ~~includes at least one of N genomic sites, N being an integer greater than or~~
17 ~~equal to one;~~

18 ~~determining a first methylation level of the first portion of the first haplotype~~
19 ~~using the first set of the plurality of DNA molecules;~~

20 ~~identifying a second set of the plurality of DNA molecules that each:~~
21 ~~is located at any one of the one or more heterozygous loci,~~
22 ~~includes the corresponding second allele of the heterozygous locus, and~~
23 ~~includes at least one of the N genomic sites;~~

24 ~~determining a second methylation level of the first portion of the second~~
25 ~~haplotype using the second set of the plurality of DNA molecules;~~

26 ~~calculating a value of a parameter using the first methylation level and the second~~
27 ~~methylation level;~~

28 ~~comparing the value of the parameter to a reference value; and~~

determining a classification of a disorder in the organism using the comparison of the value of the parameter to the reference value.

47.—The method of claim 46, wherein the first methylation level is determined using single strand methylation levels of the first set of the plurality of DNA molecules, and wherein the second methylation level is determined using single strand methylation levels of the second set of the plurality of DNA molecules.

48.—The method of claim 46, wherein the first methylation level is determined using single molecule, double stranded DNA methylation levels of the first set of the plurality of DNA molecules, and wherein the second methylation level is determined using single molecule, double stranded DNA methylation of the second set of the plurality of DNA molecules.

49.—The method of claim 46, wherein the disorder is cancer.

50.—The method of claim 46, wherein the parameter is a separation value.

51.—The method of claim 46, further comprising:
determining a plurality of first methylation levels for a plurality of portions of the first haplotype;
determining a plurality of second methylation levels for a plurality of portions of the second haplotype, each portion of plurality of portions of the second haplotype being complementary to a portion of the plurality of portions of the first haplotype,
for each portion of the plurality of portions of the second haplotype:
calculating a separation value using the second methylation level of the portion of the second haplotype and the first methylation level of the complementary portion of the first haplotype, and
comparing the separation value to a cutoff value,
wherein:
the first portion of the first haplotype is complementary to the first portion of the second haplotype, and

16 the parameter includes a number of portions of the second haplotype
17 where the separation value exceeds the cutoff value.

1 52.—The method of claim 51, wherein the cutoff value is determined from
2 tissues not having the disorder.

1 53.—The method of claim 51, wherein each portion of the plurality of portions
2 of the first haplotype has a length of greater than or equal to 5 kb.

1 54.—The method of claim 46, further comprising:
2 determining a plurality of first methylation levels for a plurality of
3 portions of the first haplotype,
4 determining a plurality of second methylation levels for a plurality of
5 portions of the second haplotype, each portion of the plurality of portions of the second
6 haplotype being complementary to a portion of the plurality of portions of the first
7 haplotype,
8 for each portion of the plurality of portions of the second haplotype:
9 calculating a separation value using the second methylation level
10 of the portion of the second haplotype and the first methylation level of the
11 complementary portion of the first haplotype,
12 wherein:
13 the first portion of the first haplotype is complementary to the first portion
14 of the second haplotype, and
15 the parameter includes a sum of the separation values.

1 55.—The method of claim 46, further comprising:
2 determining a plurality of first methylation levels for a plurality of
3 portions of the first haplotype,
4 determining a plurality of second methylation levels for a plurality of
5 portions of the second haplotype, each portion of the plurality of portions of the second
6 haplotype being complementary to a portion of the plurality of portions of the first
7 haplotype,
8 for each portion of the plurality of portions of the second haplotype:

9 calculating a separation value using the second methylation level
10 of the portion of the second haplotype and the first methylation level of the
11 complementary portion of the first haplotype, and
12 comparing the separation value to a cutoff value to identify
13 whether the portion has aberrant separation between the first methylation level
14 and the second methylation level,
15 wherein determining the classification of the disorder in the organism
16 includes comparing a pattern of portions having an aberrant separation to a reference
17 pattern.

1 56.—The method of claim 46, wherein the classification of the disorder is a
2 likelihood of the disorder.

1 57.—The method of claim 46, wherein:
2 the first portion of the first haplotype and the first portion of the second
3 haplotype form a circular DNA molecule, and
4 determining the first methylation level of the first portion of the first
5 haplotype comprises using data from the circular DNA molecule.

1 58.—The method of claim 57, wherein the circular DNA molecule is formed
2 by:
3 cutting a double stranded DNA molecule using a Cas9 complex to form a cut
4 double stranded DNA molecule, and
5 ligating a hairpin adaptor onto an end of the cut double stranded DNA molecule.

1 59.—The method of claim 46, wherein:
2 the first portion of the first haplotype is longer than or equal to 1 kb.

1 60.—The method of claim 46, wherein the reference value is determined using a
2 reference tissue without the disorder.

1 61.—The method of claim 46, wherein the disorder is an imprinting disorder.

1 62.—A method of detecting chimeric molecules in a biological sample, the
2 method comprising:
3 for each of a plurality of DNA molecules from the biological sample:
4 performing single molecule sequencing of the DNA molecule to obtain a
5 sequence read that provides a methylation status at each of N sites, N being 5 or more,
6 wherein the methylation statuses of the sequence read form a methylation pattern;
7 sliding the methylation pattern over one or more reference patterns that
8 correspond to chimeric molecules that have two portions from two parts of a reference
9 human genome, the one or more reference patterns including a change between methylated
10 statuses and unmethylated statuses; and
11 identifying a match position between the methylation pattern and a first
12 reference pattern of the one or more reference patterns, the match position identifying a
13 junction between the two parts of the reference human genome in the sequence read; and
14 outputting the junction as a location of a gene fusion in a chimeric molecule.

1 63.—The method of claim 62, wherein the match position is output to an
2 alignment function, the method further comprising:
3 refining the location of the gene fusion by:
4 aligning a first portion of the sequence read to a first part of the reference
5 human genome, the first portion being before the junction; and
6 aligning a second portion of the sequence read to a second part of the
7 reference human genome, the second portion being after the junction, wherein the first part of
8 the reference human genome is at least 1 kb apart from the second part of the human
9 reference genome.

1 64.—The method of claim 62, further comprising comparing the junctions of
2 the chimeric molecules to each other to confirm the location of the gene fusion.

1 6529. A computer product comprising a non-transitory computer readable
2 medium storing a plurality of instructions that when executed control a computer system to
3 perform the method of any one of the preceding claims.

1 6630. A system comprising:
2 the computer product of claim 6529; and
3 one or more processors for executing instructions stored on the computer readable
4 medium.

1 ~~67.—A system comprising means for performing any of the above methods.~~

1 ~~68.—A system comprising one or more processors configured to perform any of~~
2 ~~the above methods.~~

~~69.—A system comprising modules that respectively perform the steps of any~~
 ~~of the above methods.~~

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

receiving input data acquired by measuring pulses in an optical signal
 corresponding to a window of nucleotides in a sample nucleic acid molecule, and wherein the
 window comprises at least two nucleotides, the input data 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;

determining whether the methylation is present in a nucleotide at the target
 position within the window using:

for each nucleotide within the window:

the identity of the nucleotide,

the position of the nucleotide with respect to the target position,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration.

32. The method of claim 31, wherein determining whether the methylation is present in the nucleotide at the target position within the window comprises inputting the input data into a model, the model configured to:

receive the input data,

apply optimized parameters of the model to the properties of the input data; and

output a methylation status specifying whether the nucleotide at the target position in the respective window has a methylation.

1 33. The method of claim 31, wherein determining whether the methylation is
2 present in the nucleotide at the target position within the window comprises inputting the input
3 data into a model, the model trained by:

4 receiving additional data, the additional data acquired by measuring pulses in
5 optical signals corresponding to a plurality of windows of nucleotides, each window of
6 nucleotides corresponding to a first nucleic acid molecule of a plurality of first nucleic acid
7 molecules, wherein the methylation has a known first state in a nucleotide at a target position in
8 each window of each first nucleic acid molecule, the additional data comprising values for the
9 properties,

10 storing a plurality of first training samples, each including values for the
11 properties of each first nucleic acid molecule and a first label indicating the first state of the
12 nucleotide at the target position, and

13 optimizing, using the plurality of first training samples, parameters of the model
14 based on outputs of the model matching or not matching corresponding labels of the first labels
15 when the values of the properties of the first training samples are input to the model, wherein an
16 output of the model specifies whether the nucleotide at the target position in the respective
17 window has the methylation.

34. The method of claim 32 or 33, wherein the model includes a machine
learning model, a principal component analysis, a convolutional neural network, or a logistic
regression.

35. The method of any one of claims 31 to 34, wherein the values for the
properties are arranged in a matrix structure.

36. The method of any one of claims 31 to 35, wherein:

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

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

the input data comprises data acquired by measuring pulses in optical signals
corresponding to a respective window of nucleotides in each sample nucleic acid molecule of the
plurality of sample nucleic acid molecules, and

the method further comprising:

for each sample nucleic acid of the plurality of sample nucleic acid molecules,
determining whether a methylation is present in a nucleotide at a respective target location in the
respective window using:

for each nucleotide within the respective window:

the identity of the nucleotide,

the position of the nucleotide with respect to the respective target position,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration.

37. The method of claim 36, further comprising:

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

determining a classification of a disorder in a subject using the presence of the methylation at one or more nucleotides, wherein the sample nucleic acid molecule is obtained from the subject.

38. The method of claim 37, wherein the disorder comprises cancer.

39. The method of claim 37 or 38, wherein determining the classification of the disorder uses the number of methylations or the sites of the methylations.

40. The method of claim 36, further comprising:

determining a methylation status for whether the methylation is present at one or more nucleotides, and

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

41. The method of claim 40, wherein:

the method comprises determining the tissue of origin, and

determining the tissue of origin comprises determining whether the sample nucleic acid molecule is of fetal or maternal origin.

42. The method of claim 41, wherein determining whether the sample nucleic acid molecule is of fetal or maternal origin comprises:

determining a methylation level of the sample nucleic acid molecule using the methylation statuses of the one or more nucleotides, and

comparing the methylation level of the sample nucleic acid molecule to a reference value.

43. The method of any one of claims 36 to 42, further comprising filtering the biological sample for nucleic acid molecules having sizes greater than a cutoff size to obtain the plurality of sample nucleic acid molecules.

44. The method of any one of claims 36 to 43, further comprising:

aligning the plurality of sample nucleic acid molecules to a plurality of genomic regions, wherein each genomic region of the plurality of genomic regions has a subread depth greater than a cutoff number.

45. The method of any one of claims 31 to 44, further comprising sequencing the sample nucleic acid molecule.

46. The method of any one of claims 31 to 45, wherein:
the window of nucleotides comprises nucleotides on a first strand of the sample nucleic acid molecule and nucleotides on a second strand of the sample nucleic acid molecule,
and

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

47. The method of claim 46, wherein the sample nucleic acid molecule is a circular DNA molecule formed by:

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

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

48. The method of claim 31, wherein the nucleotides within the window are determined using a circular consensus sequence and are determined without alignment of the sequenced nucleotides to a reference genome.

49. The method of claim 31, wherein each nucleotide within the window is enriched or filtered.

50. The method of claim 49, wherein each nucleotide within the window is enriched by:

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

filtered by:

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

51. The method of claim 31, wherein nucleotides within the window are determined without using a circular consensus sequence and without alignment of the sequenced nucleotides to a reference genome.

52. The method of any one of claims 31 to 51, further comprising using the presence of the methylation to identify a disorder in a subject, to identify or prognosticate tumors or malignancies in the subject, to detect the tissue origin of the sample nucleic acid molecule, or

to identify chimera and hybrid DNA, wherein the sample nucleic acid molecule is obtained from the subject.

53. A computer product comprising a non-transitory computer readable medium storing a plurality of instructions that when executed control a computer system to perform the method of any one of claims 31 to 52.

54. A system comprising:
the computer product of claim 53; and
one or more processors for executing instructions stored on the computer readable medium