(d) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, the methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

(e) determining whether the sample nucleic acid molecule is of fetal or maternal origin using the methylation statuses of the one or more nucleotides.

2. The method of claim 1, 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.

3. The method of claim 2, wherein the reference value is determined from a methylation level of one or more maternal nucleic acid molecules.

4. The method of claim 2, wherein:

comparing the methylation level of the sample nucleic acid molecule to the reference value comprises determining the methylation level of the sample nucleic acid molecule is lower than the reference value, and

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

5. The method of claim 1, further comprising:

identifying the sample nucleic acid molecule as aligning to a predetermined genomic region.

6. The method of claim 1, wherein:

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

the method further comprising:

performing steps (a)-(e) for each of the plurality of sample nucleic acid molecules to determine whether each of the plurality of sample nucleic acid molecules is of fetal or maternal origin using the methylation statuses, and

Pacific Biosciences of California EXHIBIT - 1046 (Part 2)
Pacific Biosciences of California v. The Chinese U. of Hong Kong - IPR2024-00028

determining a fetal fraction using the determination of the fetal or maternal origin of the plurality of sample nucleic acid molecules.

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

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

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

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

11. The method of claim 1, wherein each window associated with each input data structure comprises at least 4 consecutive nucleotides on a first strand of each first nucleic acid molecule.

12. The method of claim 1, wherein the methylation is 5mC or 6 mA.

13. The method of claim 1, wherein:

determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, and

the sample nucleic acid molecule is a first nucleic acid molecule;

the method further comprising:

providing a second nucleic acid molecule, the second nucleic acid molecule being identical to the first nucleic acid molecule and having a methylated nucleotide at a position corresponding to the target position,

performing bisulfite sequencing on the second nucleic acid molecule to obtain bisulfite sequencing results, and

verifying the methylation present in the nucleotide at the target position using bisulfite sequencing results.

14. A method for determining whether a sample nucleic acid molecule is of fetal or maternal origin, wherein the sample nucleic acid molecule is cell-free and obtained from a biological sample of a female subject pregnant with a fetus, the method comprising:
(a) sequencing the sample nucleic acid molecule by measuring pulses in an optical signal corresponding to nucleotides in the sample nucleic acid molecule;
(b) receiving data acquired by sequencing the 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;
(c) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, 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;
(d) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, inputting the input data structure into a model, the model configured to:
receive the input data structure, generated using the data acquired by measuring the pulses in the optical signal corresponding to the nucleotides,
apply optimized parameters of the model to the properties of the input data structure; and
output a methylation status specifying whether the nucleotide at the target position in the respective window has a methylation;

(e) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, the methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

(f) determining whether the sample nucleic acid molecule is of fetal or maternal origin using the methylation statuses of the one or more nucleotides.

19. The method of claim 14, wherein sequencing the sample nucleic acid molecule comprises:

providing a polymerase that is capable of processing a first nucleic acid molecule, wherein the first nucleic acid molecule comprises the methylation in at least one nucleotide;

incorporating dye-labeled nucleotides into the first nucleic acid molecule to form the sample nucleic acid molecule; and

measuring pulses in the optical signal corresponding to the dye-labeled nucleotides, wherein determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, thereby determining the methylation is present in the at least one nucleotide in the first nucleic acid molecule.

22. The method of claim 14, further comprising:

providing a double-stranded nucleic acid molecule comprising the methylation in at least one nucleotide; and

ligating Watson and Crick strands of the double-stranded nucleic acid molecule to form a circular template molecule comprising a double-stranded segment having each end capped with a hairpin adaptor, wherein the circular template molecule is the sample nucleic acid molecule; wherein:

the data comprises sequences of the Watson and Crick strands of the double-stranded nucleic acid molecule contained in the circular template molecule, and

determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample

nucleic acid molecule at the target position, thereby determining the methylation is present in the at least one nucleotide in the double-stranded nucleic acid molecule.

23. The method of claim 14, further comprising aligning the sequenced nucleotides to a reference genome to provide the positions of the nucleotides within the sample nucleic acid molecule.

3. Claims 1, 2, 6-8, 11, 12, 14-16, 21-25, 27, 28, 31-43 are rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1-19 of **U.S. Patent No. 11,091,794.** Although the claims at issue are not identical, they are not patentably distinct from each other because the US Patent and instant claims are both drawn to overlapping methods of detecting methylation of nucleotides encompassing cytosine in a nucleic acid molecule and utilizing a machine learning program to do so, rendering the instant computer system conducting the method obvious. The US Patent claims:

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.

2. The method of claim 1, wherein:
the input data structure is one input data structure of a plurality of input data structures,
the sample nucleic acid molecule is one sample nucleic acid molecule of a plurality of sample nucleic acid molecules,

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

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

the method further comprising:

receiving the plurality of input data structures,

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

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

3. The method of claim 2, further comprising:

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

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

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

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

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

the method further comprising:

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

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

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

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

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

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

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

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

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

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

16. The method of claim 15, wherein each nucleotide within the window is enriched by: cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule, or
filtered by:
selecting double-stranded DNA molecules having a size with a size range.

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

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

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

4. **Conclusion:** No claims are allowed. The closest prior art made of record but not relied upon are Summerer (ChemBioChem, 2010, 11:2499-2501); US Patent 9,175,348; US Patent 9,175,338; and US Patent 10,590,484. Summerer and all three US Patents teach methods for single molecule real time (SMRT) sequencing to identify cytosine methylation in sequences. The US Patents teach measuring pulse width of nucleotides, interpulse duration (IPD), and aligning sequences and signals to identify differences in IPD and pulse width as indicators of genetic alteration such as methylation (Figure 2 in Summerer, US Patent 9,175,348 and US Patent 9,175,338). None of Summerer and the

US Patents demonstrate measuring a pulse width value of a specifically identified nucleotide at an identified position to determine the methylation status, and do not teach or suggest combining all data measurements of nucleotide identities, nucleotide positions in the sample nucleic acid, width of the pulse corresponding to the nucleotides, and IPD, subsequently creating an input data structure comprising a window of the nucleotides in a sample that lists the above measurements, and inputting the measurements into a model configured to utilize the measurements to output a methylation status for each nucleotide as instantly claimed.

5. Any inquiry concerning this communication or earlier communications from the examiner should be directed to LAURA B GODDARD whose telephone number is (571)272-8788. The examiner can normally be reached Mon-Fri, 7am-3:30pm.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Misook Yu can be reached on 571-272-0839. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of published or unpublished applications may be obtained from Patent Center. Unpublished application information in Patent Center is available to registered users. To file and manage patent submissions in Patent Center, visit: https://patentcenter.uspto.gov. Visit https://www.uspto.gov/patents/apply/patent-

center for more information about Patent Center and

https://www.uspto.gov/patents/docx for information about filing in DOCX format. For

additional questions, contact the Electronic Business Center (EBC) at 866-217-9197

(toll-free). If you would like assistance from a USPTO Customer Service

Representative, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Laura B Goddard/
Primary Examiner, Art Unit 1642

Notice of References Cited	Application/Control No. 17/891,899	Applicant(s)/Patent Under Reexamination Lo et al.	
	Examiner LAURA B GODDARD	Art Unit 1642	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-9175348-B2	11-2015	Korlach; Jonas	C12Q1/6883	1/1
*	B	US-9175338-B2	11-2015	Flusberg; Benjamin	C12Q1/6869	1/1
*	C	US-10590484-B2	03-2020	Korlach; Jonas	C12Q1/6874	1/1
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Summerer (ChemBioChem, 2010, 11:2499-2501)
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/891,899	Lo et al.
	Examiner	Art Unit
	LAURA B GODDARD	1642

CPC - Searched*		
Symbol	Date	Examiner

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
STN	04/11/2023	LBG
PE2E search	04/11/2023	LBG

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

/Laura B Goddard/ Primary Examiner, Art Unit 1642	

* * * * * * * * * * Welcome to STN International * * * * * * * * * *

NEWS 1 MAR 25 New Get Prior Art Analysis Feature in CA/CAplus Provides an Efficient Means to Expand Search Results
NEWS 2 MAR 25 Searching Chemical Compounds with Fragmentation Codes in Derwent World Patents Index (R) Subscriber Files
NEWS 3 MAR 30 New Projects Feature Allows STNext Users to Define and Manage Search Projects
NEWS 4 APR 4 The enhanced Derwent World Patents Index(R) is now available on STNext
NEWS 5 APR 4 The Derwent Chemistry Resource (DCR) is now available as a standalone database on STNext
NEWS 6 JUN 22 Claims for Four Additional Patent Authorities Added to CA/CAplus
NEWS 7 JUN 24 SMARTracker SDI Option Now Supported in Derwent Files
NEWS 8 JUN 24 Expanded Claims Coverage and Enhanced HIT Displays of Claims Data in Derwent World Patents Index(R)
NEWS 9 SEP 9 Enhanced Coverage of Malaysia, Saudi Arabia, and Georgia Increases Comprehensiveness of INPADOC
NEWS 10 SEP 14 New Preference Setting for Transcript Options Helps Better Manage Transcripts, Increase Their Efficiency
NEWS 11 SEP 22 CAS STNext Structure Editor Update Preserves Functionality, Increases Flexibility
NEWS 12 OCT 24 Links to National Patent Registers Now Available in Patent Databases on CAS STNext
NEWS 13 NOV 7 PCTFULL Reload Provides Additional Coverage, Independent Claims and Assignee Searching Enhancements
NEWS 14 NOV 7 EPFULL Reload on STNext Features New Independent Claims Searching Capabilities
NEWS 15 NOV 10 PS (Pharmaceutical Substances) Database Reloaded
NEWS 16 DEC 05 Expanded access to Taiwanese patent information provided by new full text database TWFULL and claims coverage in CAplus(SM)
NEWS 17 DEC 05 New Implementation of Derwent Drug File (DRUGU/DDFU) Makes Searching More Convenient
NEWS 18 DEC 23 New Interactive Claim Viewer in EPFULL and PCTFULL Makes Claim Analysis More Efficient
NEWS 19 DEC 30 2023 MeSH now available in MEDLINE on STNext
NEWS 20 FEB 03 January 2023 Update to Emtree(R) Now Available in Embase(R)
NEWS 21 FEB 17 Identification of Substances from Claims Simplified in CAS PatentPak - UPDATE
NEWS 22 FEB 23 Derwent World Patents Index - Manual Code revision 2023
NEWS 23 MAR 10 MEDLINE(R) Reloaded and MeSH Thesaurus Updated for 2023 on STNext
NEWS 24 MAR 23 Expanded Patent Authority Coverage in CA/CAplus Means Access to More Inventions, More Complete Patent Families

NEWS HOURS STN Operating Hours Plus Help Desk Availability
NEWS LOGIN Welcome Banner and News Items
NEWS TRAINING Find instructor-led and self-directed training opportunities

Enter NEWS followed by the item number or name to see news on that specific topic.

For pricing information enter HELP PRICE or HELP COST at any arrow prompt.

All use of STN is subject to the provisions of the STN customer agreement. This agreement limits use to scientific research. Use for software development or design, implementation of commercial gateways, or use of CAS and STN data in the building of commercial products is prohibited and may result in loss of user privileges and other penalties.

Access the latest version of STN.
Log into STNext with your STN login ID at https://www.stn.org

STNext is updated frequently. To see the latest STNext Release Notes, click

https://www.stn.org/help/stnext/ReleaseNotes.htm

```
* * * * * * * * * * * * * * * * STN  Columbus  * * * * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 09:48:38 ON 11 APR 2023

=> FILE .BIO

COST IN U.S. DOLLARS                            SINCE FILE            TOTAL
                                                     ENTRY          SESSION
FULL ESTIMATED COST                                   1.40             1.40

FILE 'CAPLUS' ENTERED AT 09:51:48 ON 11 APR 2023
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
COPYRIGHT (C) 2023 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'EMBASE' ENTERED AT 09:51:48 ON 11 APR 2023
Copyright (c) 2023 Elsevier B.V. All rights reserved.

FILE 'BIOSIS' ENTERED AT 09:51:48 ON 11 APR 2023
Copyright (c) 2023 Clarivate Analytics

FILE 'MEDLINE' ENTERED AT 09:51:48 ON 11 APR 2023

FILE 'SCISEARCH' ENTERED AT 09:51:48 ON 11 APR 2023
Copyright (c) 2023 Clarivate Analytics

=> S ("SINGLE MOLECULE REAL TIME SEQUENCING" OR SMRT) AND (PULSE OR PEAK)(S)(WIDTH
OR DISTANCE)

L1          11 ("SINGLE MOLECULE REAL TIME SEQUENCING" OR SMRT) AND (PULSE OR
               PEAK)(S)(WIDTH OR DISTANCE)

=> S L1 AND cytosine AND (methylation OR methylated)

L2           6 L1 AND CYTOSINE AND (METHYLATION OR METHYLATED)

=> S L2 AND py<2020

L3           1 L2 AND PY<2020

=> DUP REM L3

PROCESSING COMPLETED FOR L3
L4           1 DUP REM L3 (0 DUPLICATES REMOVED)

=> D L4 IBIB ABS HITINF

'HITINF' IS NOT A VALID FORMAT FOR FILE 'CAPLUS'

For a complete list of valid display formats and field codes,
enter HELP FORMAT or HELP DFIELDS at an arrow prompt (=>).

If you need further assistance, contact your local Help Desk.

TYPE "END" TO RETURN TO AN ARROW PROMPT OR
ENTER DISPLAY FORMAT (BIB):IBIB ABS HITIND


L4   ANSWER 1 OF 1  CAPLUS  COPYRIGHT 2023 ACS on STN
ACCESSION NUMBER:        2010:1522246  CAPLUS Full-text
DOCUMENT NUMBER:         154:125720
TITLE:                   High-Throughput DNA Sequencing Beyond the Four-Letter
                         Code: Epigenetic Modifications Revealed by
                         Single-Molecule Bypass Kinetics
AUTHOR(S):               Summerer, Daniel
CORPORATE SOURCE:        Febit Biomed GmbH, Heidelberg, 69120, Germany
SOURCE:                  ChemBioChem (2010), 11(18), 2499-2501
                         CODEN: CBCHFX; ISSN: 1439-4227
DIGITAL OBJECT ID:       10.1002/cbic.201000569
PUBLISHER:               Wiley-VCH Verlag GmbH & Co. KGaA
```

DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review and discussion of the single-mol., real-time (SMRT) DNA sequencing procedure developed by B. A. Flusberg et al. (2010) which allows for the direct, bisulfite-free detection of methylation in DNA sequences. SMRT sequencing uses zeptoliter-sized nanostructures called zero-mode waveguides (ZMW) that permit fluorescence detection with minimized background. Single DNA polymerases can be immobilized in ZMWs and can perform processive primer extension of single-sample DNA mols. Using dNTP with specific fluorecent labels on the γ-phosphate, template-specific incorporation of a dNTP results in temporal localization of the label in the ZMW and subsequent release after pyrophosphate dissocn. Besides the order of dNTP incorporation, 2 addnl. kinetic parameters can be extd. from the obtained fluorescence traces: the pulse width, which is a function of all the kinetic steps of dNTP incorporation until pyrophosphate release, and the interpulse duration (IPD), which is dependent on the dNTP binding and DNA polymerase translocation. The IPD is sensitive to epigenetic modifications in the template strand and can be used for detection of N-6-methyladenine, and potentially 5-methylcytosine and 5-hydroxymethylcytosine, in untreated DNA samples.
CC 3-0 (Biochemical Genetics)
Section cross-reference(s): 9
ST review methyl adenine cytosine DNA sequencing nanostructure; single mol bypass kinetics epigenetic modification DNA sequencing review
IT Methylation
(of DNA; nanostructure-based DNA sequencing method detects epigenetic modifications by single-mol. bypass kinetics)
OS.CITING REF COUNT: 7 THERE ARE 7 CAPLUS RECORDS THAT CITE THIS RECORD (7 CITINGS)
REFERENCE COUNT: 24 THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

=>

=>

PE2E SEARCH - Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	British Equivalents	Time Stamp
L1	1384	("single molecule real time sequencing" SMRT) AND (methylation methylated) AND (pulse interpulse)	(US-PGPUB; USPAT; USOCR; FIT (AU, AP, AT, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, OA, RU, SU, WO, MC, MD, MY, NL, NO, NZ, PH, PL, PT, RO, RS, SE, SG, SI, SK, TH, TN, TR, TW, UA, VN); FPRS; EPO; JPO; DERWENT)	OR	ON	ON	2023/04/10 02:08 PM
L2	351	L1 AND ((width duration) NEAR (pulse))	(US-PGPUB; USPAT; USOCR; FIT (AU, AP, AT, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, OA, RU, SU, WO, MC, MD, MY, NL, NO, NZ, PH, PL, PT, RO, RS, SE, SG, SI, SK, TH, TN, TR, TW, UA, VN); FPRS; EPO; JPO; DERWENT)	OR	ON	ON	2023/04/10 02:08 PM
L3	22	L2 AND "chinese university of hong kong"	(US-PGPUB; USPAT; USOCR; FIT (AU, AP, AT, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, OA, RU, SU, WO, MC, MD, MY, NL, NO, NZ, PH, PL, PT, RO, RS, SE, SG, SI, SK, TH, TN, TR, TW, UA, VN); FPRS; EPO; JPO; DERWENT)	OR	ON	ON	2023/04/10 02:10 PM
L4	248	L2 AND @ay<="2019"	(US-PGPUB; USPAT; USOCR; FIT (AP, AT, AU, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS,	OR	ON	ON	2023/04/10 02:15 PM

			IT, JP, KR, LT, LU, LV, MA, MC, MD, MY, NL, NO, NZ, OA, PH, PL, PT, RO, RS, RU, SE, SG, SI, SK, SU, TH, TN, TR, TW, UA, VN, WO); FPRS; EPO; JPO; DERWENT)				
L5	14	L4 AND ((width duration) AND (pulse) AND (methylation methylated)).clm.	(US-PGPUB; USPAT; USOCR; FIT (AP, AT, AU, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, MC, MD, MY, NL, NO, NZ, OA, PH, PL, PT, RO, RS, RU, SE, SG, SI, SK, SU, TH, TN, TR, TW, UA, VN, WO); FPRS; EPO; JPO; DERWENT)	OR	ON	ON	2023/04/10 02:16 PM
L6	4	L5 AND (matrix grid table window).clm.	(US-PGPUB; USPAT; USOCR; FIT (AP, AT, AU, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, MC, MD, MY, NL, NO, NZ, OA, PH, PL, PT, RO, RS, RU, SE, SG, SI, SK, SU, TH, TN, TR, TW, UA, VN, WO); FPRS; EPO; JPO; DERWENT)	OR	ON	ON	2023/04/10 02:18 PM
L7	38	L2 AND ((width duration) AND (pulse) AND (methylation methylated)).clm.	(US-PGPUB; USPAT; USOCR; FIT (AP, AT, AU, BE, BG, BR, BY, CA, CH, CN, CS, CU, CZ, DD, DE, DK, EA, EE, EP, ES, FI, FR, GB, HR, HU, ID, IE, IL, IS, IT, JP, KR, LT, LU, LV, MA, MC, MD, MY, NL, NO, NZ, OA, PH, PL, PT, RO, RS, RU, SE, SG, SI, SK, SU, TH, TN, TR, TW, UA, VN, WO); FPRS; EPO; JPO; DERWENT)	OR	ON	ON	2023/04/10 02:22 PM

There are no Interference searches to show.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

U.S. PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code [1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1	8467976	B2	Jun 18, 2013	Lo et al.	
	2	9116118	B2	Aug 25, 2015	Turner et al.	
	3	9175338	B2	Nov 3, 2015	Flusberg et al.	
	4	9175341	B2	Nov 3, 2015	Flusberg et al.	
	5	9175348	B2	Nov 3, 2015	Korlach et al.	
	6	10438691	B2	Oct 8, 2019	Kim et al.	
	7	10465232	B1	Nov 5, 2019	Wu et al.	
	8	10468121	B2	Nov 5, 2019	Kermani et al.	
	9	10590484	B2	Mar 17, 2020	Korlach et al.	

U.S. PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code [1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	10	20070042365	A1	Feb 22, 2007	Millar et al.	
	11	20100221716	A1	Sep 2, 2010	Flusberg et al.	
	12	20110183320	A1	Jul 28, 2011	Flusberg et al.	
	13	20130230909	A1	Sep 5, 2013	Pan et al.	
	14	20140004511	A1	Jan 2, 2014	Korlach et al.	
	15	20160017419	A1	Jan 21, 2016	Chiu et al.	
	16	20170037464	A1	Feb 9, 2017	Turner et al.	
	17	20170159132	A1	Jun 8, 2017	Okino et al.	
	18	20170233802	A1	Aug 17, 2017	Flusberg et al.	
	19	20180201993	A1	Jul 19, 2018	Turner et al.	
	20	20190024162	A1	Jan 24, 2019	Korlach et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	17/891,899
Filing Date	August 19, 2022
First Named Inventor	Yuk-Ming Dennis Lo
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	080015-1343271

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T[5]
	21	2002214811	AU	B2	Sep 9, 2004	Human Genetic Signatures Pty Ltd.		
	22	2005024053	WO	A1	Mar 17, 2005	Human Genetic Signatures Pty Ltd.		
	23	2009013492	WO	A1	Jan 29, 2009	The Chinese University of Hong Kong		
	24	2010027484	WO	A2	Mar 11, 2010	Pacific Biosciences of California, Inc.		
	25	2010068289	WO	A2	Jun 17, 2010	Pacific Biosciences of California, Inc.		
	26	2012170429	WO	A1	Dec 13, 2012	United States Postal Service		
	27	2013170429	WO	A1	Nov 21, 2013	BGI Genomics Co., Ltd.	English Abstract Attached	
	28	2013188846	WO	A1	Dec 19, 2013	Harry Stylli		
	29	104053789	CN	A	Sep 17, 2014	BGI Health Service Co., Ltd.	English Abstract Attached	
	30	2014153757	WO	A1	Oct 2, 2014	BGI Shenzhen	English Abstract Attached	
	31	104053789	CN	B	Feb 10, 2016	BGI Genomics Co., Ltd.	English Abstract Attached	
	32	2012380221	AU	B2	Sep 29, 2016	BGI Genomics Co., Ltd.		
	33	2017012544	WO	A1	Jan 26, 2017	The Chinese University of Hong Kong		
	34	101896167	KR	B1	Sep 7, 2018	The Chinese University of Hong Kong	English Abstract of corresponding document: W02009013492 (A1) Attached	

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	35	Machine Learning, Wikipedia, Available Online at: https://en.wikipedia.org/w/index.php?title=Machine learning&oldid=910906899#Free and open-source software, Assessed on internet at: October 12, 2021, 21 pages.	
	36	U.S. Patent Application No. 16/995,607, Non-Final Office Action mailed on December 30, 2020, 19 pages.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

	37	U.S. Patent Application No. 17/379,544, Non-Final Office Action mailed on February 4, 2022, 10 pages.	
	38	Australian Application No. 2020323958, First Examination Report mailed on May 31, 2021, 3 pages.	
	39	Australian Application No. 2022231719, First Examination Report mailed on September 28, 2022, 4 pages.	
	40	BEAULAURIER et al., Deciphering Bacterial Epigenomes Using Modern Sequencing Technologies, Nature Reviews Genetics, Volume 20 Issue 3, March 2019, pp. 157-172.	
	41	BLOW et al., The Epigenomic Landscape of Prokaryotes, PLOS Genetics, Volume 12, No. 2, February 12, 2016, pp. 1-28.	
	42	CLARK et al., Characterization of DNA Methyltransferase Specificities Using Single-Molecule, Real-Time DNA Sequencing, Nucleic Acids Research, Volume 40, No. 4, Available online at: https://doi.org/10.1093/nar/gkr1146, 2012, 12 pages	
	43	CLARK et al., Enhanced 5-Methylcytosine Detection in Single-Molecule, Real-Time Sequencing via Tet1 Oxidation, BMC Biology, Volume 11, No. 4, 2013, 10 pages.	
	44	EID et al., Real-Time DNA Sequencing from Single Polymerase Molecules, Science, Volume 323, No. 5910, January 2, 2009, pp. 133-138.	
	45	European Application No. 20853612.8, Extended European Search Report Action mailed on October 25, 2021, 15 pages.	
	46	FENG et al., Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic, PLOS Computational Biology, Volume 9, No. 3, Available online at: https://doi.org/10.1371/journal.pcbi.1002935, March 2013, 10 pages.	
	47	FENG et al., qDNAmod: A Statistical Model-based Tool to Reveal Intercellular Heterogeneity of DNA Modification from SMRT Sequencing Data, Nucleic Acids Research, Volume 42, No. 22, November 17, 2014, pp. 13488-13499.	
	48	FLUSBERG et al., Direct Detection of DNA Methylation During Single-Molecule, Real-Time Sequencing, Nature Methods, Volume 7, No. 6, Available Online at: htt12s://12ubmed.ncbi.nlm.nih.gov/20453 866/, June 2010, pp. 461-465,	
	49	United Kingdom Application No. 2102808.9, First Examination Report mailed on April 12, 2021, 6 pages.	
	50	United Kingdom Application No. 2102808.9, Office Action mailed on September 10, 2021, 4 pages.	
	51	United Kingdom Application No. 2102808.9, Office Action mailed on June 23, 2021, 8 pages.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

	52	United Kingdom Application No. 2103003.6, Combined Search and Examination Report mailed on March 29, 2021, 5 pages.	
	53	United Kingdom Application No. 2103010.1, Combined Search and Examination Report mailed on April 15, 2021, 4 pages.	
	54	United Kingdom Application No. 2201811.3, Combined Search and Examination Report mailed on March 1, 2022, 4 pages.	
	55	United Kingdom Application No. 2201821.2, Search Report mailed on February 24, 2022, 4 pages.	
	56	United Kingdom Application No. 2210631.4, Search Report mailed on November 7, 2022, 4 pages.	
	57	United Kingdom Application No. 2210633.0, Search Report mailed on August 26, 2022, 4 pages.	
	58	United Kingdom Application No. 2216178.0, Search Report mailed on November 15, 2022, 4 pages.	
	59	HE et al., TETI Promotes Leukemic Growth of AMLI-ETO+ AML Via 5-Hydroxymethylcytosine Marks and its Oncogenic Role and High Expression can be Antagonized Using Olaparib, an Inhibitor of its Binding Partner PARPI, Experimental Hematology, Volume 53, 2017, pp. S92-S93.	
	60	Japanese Application No. 2021-514525, Office Action mailed on July 26, 2022, 7 pages (4 pages of English Translation and 3 pages of Original Document).	
	61	KILE et al., Correlation of Global and Gene-Specific DNA Methylation in Maternal-Infant Pairs, PLoS One, Volume 5, No. 10, October 29, 2010, 9 pages.	
	62	LIU et al., Bisulfite-Free Direct Detection of 5-Methylcytosine and 5- Hydroxymethylcytosine at Base Resolution, Nature Biotechnology Letters, Volume 37, April 2019, pp. 424-429.	
	63	LIU et al., Detection of DNA Base Modifications by Deep Recurrent Neural Network on Oxford Nanopore Sequencing Data, Nature Communications, Volume 10, No. 2449, June 4, 2019, 11 pages.	
	64	NI et al., Deepsignal: Detecting DNA Methylation State from Nanopore Sequencing Reads Using Deep-Learning, Bioinformatics, Volume 35, No. 22, 2019, pp. 4586-4595.	
	65	International Application No. PCT/CN2020/109602, International Search Report and Written Opinion mailed on November 19, 2020, 11 pages.	
	66	SCHADT et al., Modeling Kinetic Rate Variation in Third Generation DNA Sequencing Data to Detect Putative Modifications to DNA Bases, Genome Research, Volume 23, No. 1, January 2013, pp. 129-141.	
	67	Singapore Application No. 11202101070Q, Written Opinion mailed on October 28, 2021, 9 pages.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

	68	SUGAI et al., Analysis of the DNA Methylation Level of Cancer-related Genes in Colorectal Cancer and the Surrounding Normal Mucosa, Clinical Epigenetics, Volume 9, No. 55, May 18, 2017, pp. 1-10.	
	69	SUMMERER et al., High-Throughput DNA Sequencing Beyond the Four-Letter Code: Epigenetic Modifications Revealed by Single-Molecule Bypass Kinetics, ChemBioChem, Volume 11, No. 8, 2010, pp. 2499-2501.	
	70	SUZUKI et al., Agin: Measuring the Landscape of CpG Methylation of Individual Repetitive Elements, Bioinformatics, Volume 32, No.19, June 17, 2016, pp. 2911-2919.	
	71	SUZUKI, Observing Comprehensive DNA Methylomes via Single Molecule Real-Time Sequencing: Application to Diploid and Centromeric Methylation, Dissertation submitted to The Graduate School of Frontier Sciences, The University of Tokyo, December 13, 2017, 118 pages.	

EXAMINER SIGNATURE

Examiner Signature	/Laura B Goddard/	Date Considered	04/11/2023

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document.
[4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David D. Hsu/	Date (YYYY-MM-DD)	2023-02-02
Name/Print	David D. Hsu	Registration Number	71659

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/891,899	08/19/2022	Yuk-Ming Dennis Lo	080015-028430US-1343271	1331

20350 7590 03/21/2023

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

EXAMINER
CENTRAL, DOCKET

ART UNIT	PAPER NUMBER
1642	

NOTIFICATION DATE	DELIVERY MODE
03/21/2023	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

KTSDocketing2@kilpatrick.foundationip.com
ipefiling@kilpatricktownsend.com

Decision Granting Request for Prioritized Examination (Track I)	**Application No.** 17/891,899	**Applicant(s)** Lo et al.	
	Examiner IRVIN DINGLE	**Art Unit** OPET	**AIA (FITF) Status** Yes

1. THE REQUEST FILED 19 August 2022 IS **GRANTED** .

The above-identified application has met the requirements for prioritized examination

A. ☑ for an original nonprovisional application (Track I).
B. ☐ for an application undergoing continued examination (RCE).

2. **The above-identified application will undergo prioritized examination**. The application will be accorded special status throughout its entire course of prosecution until one of the following occurs:

A. filing a **petition for extension of time** to extend the time period for filing a reply;

B. filing an **amendment to amend the application to contain more than four independent claims, more than thirty total claims** , or a multiple dependent claim;

C. filing a **request for continued examination** ;

D. filing a notice of appeal;

E. filing a request for suspension of action;

F. mailing of a notice of allowance;

G. mailing of a final Office action;

H. completion of examination as defined in 37 CFR 41.102; or

I. abandonment of the application.

Telephone inquiries with regard to this decision should be directed to IRVIN DINGLE at (571)272-3210. In his/her absence, calls may be directed to Petition Help Desk at (571) 272-3282.

/IRVIN DINGLE/ Lead Paralegal Specialist, OPET	

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
17/891,899	08/19/2022		4080	080015-028430US-1343271	30	3

CONFIRMATION NO. 1331

UPDATED FILING RECEIPT

20350
Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

Date Mailed: 03/08/2023

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Yuk-Ming Dennis Lo, Kowloon, CHINA;
Rossa Wai Kwun Chiu, Shatin, CHINA;
Kwan Chee Chan, Kowloon, CHINA;
Peiyong Jiang, Tai Po, CHINA;
Suk Hang Cheng, Fanling, CHINA;
Wenlei Peng, Shatin, CHINA;
On Yee Tse, Fanling, CHINA;

Applicant(s)

The Chinese University of Hong Kong, Shatin, HONG KONG;

Power of Attorney: The patent practitioners associated with Customer Number 20350

Domestic Priority data as claimed by applicant

This application is a CON of 17/379,544 07/19/2021 PAT 11,466,308
which is a CON of 16/995,607 08/17/2020 PAT 11,091,794
which claims benefit of 63/051,210 07/13/2020
and claims benefit of 63/019,790 05/04/2020
and claims benefit of 62/991,891 03/19/2020
and claims benefit of 62/970,586 02/05/2020
and claims benefit of 62/887,987 08/16/2019

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.

Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 02/13/2023

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 17/891,899**

Projected Publication Date: 06/15/2023

Non-Publication Request: No

Early Publication Request: No
Title

DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS

Preliminary Class

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign

patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 17/891,899

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A			N/A	320
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A			N/A	700
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	800
TOTAL CLAIMS (37 CFR 1.16(i))	30 minus 20 =	* 10			OR	x 100 =	1000
INDEPENDENT CLAIMS (37 CFR 1.16(h))	3 minus 3 =	*				x =	0
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).						1260
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))							0
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL			TOTAL	3080

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)		RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

I hereby certify that this correspondence is being filed with the United States Patent and Trademark Office via the Office electronic filing system on March 6, 2023.

KILPATRICK TOWNSEND & STOCKTON LLP

By: /Heather A. Orosco/
Heather A. Orosco

PATENT
Attorney Docket No.: 080015-028430US-1343271
Client Ref. No.: 19/MED/901 US Con2

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Yuk-Ming Dennis Lo et al. Application No.: 17/891,899 Filed: August 19, 2022 For: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS Customer No.: 20350	Confirmation No.: 1331 Examiner: Not Yet Known Technology Center/Art Unit: Not Yet Known **PRELIMINARY AMENDMENT AND RESPONSE TO NOTICE TO FILE CORRECTED APPLICATION PAPERS**

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Commissioner:

Pursuant to the *Notice to File Corrected Application Papers* of Nonprovisional Application, dated February 14, 2023, transmitted herewith are papers in the above-identified application.

☒ Formalities Letter – Notice to File Corrected Application Papers – Filing Date Granted.

☒ Replacement Drawing Sheet for FIG. 133.

☐ Petition for Extension of Time.

☒ The Commissioner is hereby authorized to charge any additional fees required under 37 CFR §1.16 to Account No. 20-1430.

In addition, prior to examination of the application, please enter the following amendments and consider the following remarks:

Amendments to the Drawings begin on page 2 of this paper and includes an attached Replacement Sheet (and an annotated sheet showing changes, if applicable).

Remarks/Arguments begin on page 3 of this paper.

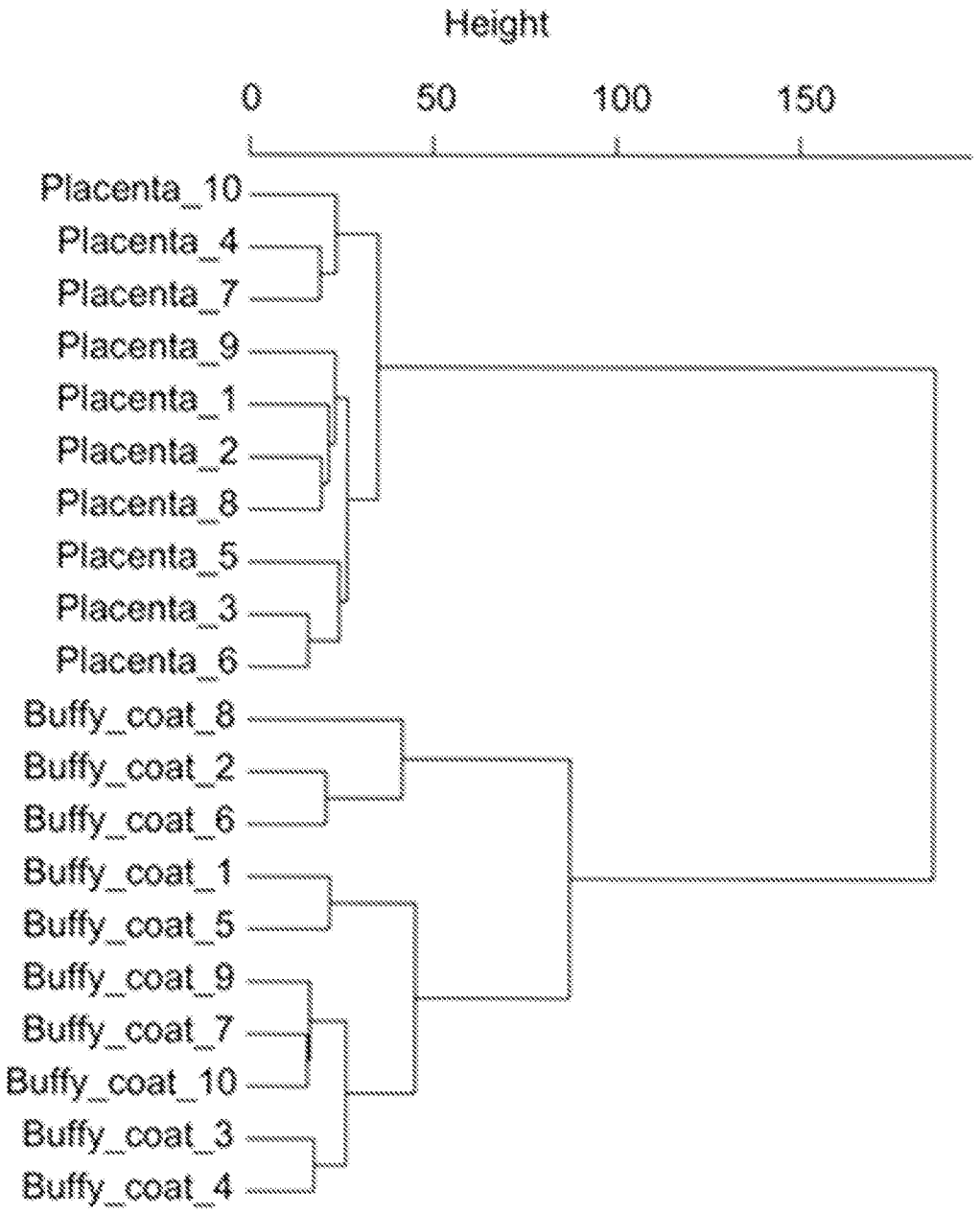

Height
0
50
100
150
Placenta_10
Placenta_4
Placenta_7
Placenta_9
Placenta_1
Placenta_2
Placenta_8
Placenta_5
Placenta_3
Placenta_6
Buffy_coat_8
Buffy_coat_2
Buffy_coat_6
Buffy_coat_1
Buffy_coat_5
Buffy_coat_9
Buffy_coat_7
Buffy_coat_10
Buffy_coat_3
Buffy_coat_4

FIG. 133

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
17/891,899	08/19/2022	Yuk-Ming Dennis Lo	080015-028430US-1343271

CONFIRMATION NO. 1331

20350
Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

FORMALITIES LETTER

OC000000053746505

Date Mailed: 02/14/2023

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Filing Date Granted

An application number and filing date have been accorded to the above-identified nonprovisional application. The application is informal since it does not comply with the regulations for the reason(s) indicated below. Applicant is given **TWO MONTHS** from the date of this Notice within which to file all required items below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- **Drawing Items**

  The drawings submitted on 08/19/2022 are not acceptable because:
  - Figure(s) 133 contain figure or view numbers that have incorrect orientation. Reference characters, sheet numbers, and view numbers must be oriented in the same direction as the view. See 37 CFR 1.84(p)(1). Replacement drawing sheets in compliance with 37 CFR 1.84 and 1.121(d) having all labeling oriented in the same direction are required.

Replies must be received in the USPTO within the set time period or must include a proper Certificate of Mailing or Transmission under 37 CFR 1.8 with a mailing or transmission date within the set time period. For more information and a suggested format, see Form PTO/SB/92 and MPEP 512.

Replies should be addressed to:

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

REPLY VIA ELECTRONIC FILING

A reply may be submitted via electronic filing only by registered users of the appropriate USPTO patent electronic filing system and must include:

- A separate document having the document description, "**Applicant response to Pre-Exam Formalities Notice**"; and
- the item(s) for which a reply period is set forth above.

For more information about electronic filing, contact the Patent Electronic Business Center (EBC) 1-866-217-9197 or visit https://www.uspto.gov/ebc

NOTE: If the reply is not being submitted electronically, a copy of this notice must be included.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/hchin/

Electronic Acknowledgement Receipt

EFS ID:	47635055
Application Number:	17891899
International Application Number:	
Confirmation Number:	1331
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Customer Number:	20350
Filer:	David Dah-Wei Hsu/Heather Orosco
Filer Authorized By:	David Dah-Wei Hsu
Attorney Docket Number:	080015-028430US-1343271
Receipt Date:	06-MAR-2023
Filing Date:	19-AUG-2022
Time Stamp:	14:55:10
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		1343271_Preliminary_Amendment.pdf	127518 3478b07206ec8058807986a31ab032b20ea9aa3e	yes	3

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Preliminary Amendment	1	1
	Drawings-only black and white line drawings	2	2
	Applicant Arguments/Remarks Made in an Amendment	3	3

Warnings:					
Information:					
2	Drawings-only black and white line drawings	1343271_Replacement_Sheet_for_FIG_133.pdf	231100 944ce5acc74cf84179d17011c6455c12124e631e	no	1
Warnings:					
Information:					
3	Miscellaneous Incoming Letter	1343271_Corrected_Application_Papers_Notice.pdf	63707 23afcaa5c1a9eb750dd650d8f39efff78a8d94b7	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	422325		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

AMENDMENT(S) TO THE DRAWINGS

The attached sheet of drawings includes changes to FIG. 133. This sheet, which includes FIG. 133, replaces the original sheet including FIG. 133.

Attachment(s): Replacement Sheet

REMARKS/ARGUMENTS

The replacement drawing sheet for FIG. 133 is submitted in response to the Notice to File Corrected Application Papers mailed on February 14, 2023. FIG. 133 has been amended so that all labeling is oriented in the same direction as the view. No new matter has been added.

Applicant respectfully requests entry of the replacement drawing sheet and submits that a corresponding amendment to the specification is not necessary for the replacement drawing sheet. Due to the simplicity of the amendments, an annotated sheet showing the changes is not included.

CONCLUSION

In view of the foregoing, Applicant believes all claims now pending in this application are in condition for examination.

Except for the issue fees payable under 37 C.F.R. § 1.18, the Director is authorized to charge any additional fees during pendency of this application, including any required extension of time fees, or credit any overpayment to Deposit Account Number 20-1430. This paragraph is intended to be a constructive petition for extension of time in accordance with 37 C.F.R. § 1.136(a)(3).

If the Examiner believes a telephone conference would expedite prosecution of this application, please contact the undersigned at (303) 571-4000 or dhsu@kilpatricktownsend.com.

Respectfully submitted,

/David D. Hsu/
David D. Hsu
Registration No. 71659

KILPATRICK TOWNSEND & STOCKTON LLP

DDH:hao

77102446V.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
17/891,899	08/19/2022		4080	080015-028430US-1343271	30	3

CONFIRMATION NO. 1331

FILING RECEIPT

20350
Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

Date Mailed: 02/14/2023

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Yuk-Ming Dennis Lo, Kowloon, CHINA;
Rossa Wai Kwun Chiu, Shatin, CHINA;
Kwan Chee Chan, Kowloon, CHINA;
Peiyong Jiang, Tai Po, CHINA;
Suk Hang Cheng, Fanling, CHINA;
Wenlei Peng, Shatin, CHINA;
On Yee Tse, Fanling, CHINA;

Applicant(s)

The Chinese University of Hong Kong, Shatin, HONG KONG;

Power of Attorney: The patent practitioners associated with Customer Number 20350

Domestic Priority data as claimed by applicant

This application is a CON of 17/379,544 07/19/2021 PAT 11,466,308
which is a CON of 16/995,607 08/17/2020 PAT 11,091,794
which claims benefit of 63/051,210 07/13/2020
and claims benefit of 63/019,790 05/04/2020
and claims benefit of 62/991,891 03/19/2020
and claims benefit of 62/970,586 02/05/2020
and claims benefit of 62/887,987 08/16/2019

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.

Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 02/13/2023

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 17/891,899**

Projected Publication Date: To Be Determined - pending completion of Corrected Papers.

Non-Publication Request: No

Early Publication Request: No
Title

DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS

Preliminary Class

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign

patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 17/891,899

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A			N/A	320
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A			N/A	700
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	800
TOTAL CLAIMS (37 CFR 1.16(i))	30 minus 20 =	* 10			OR	x 100 =	1000
INDEPENDENT CLAIMS (37 CFR 1.16(h))	3 minus 3 =	*				x =	0
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).						1260
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))							0
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL			TOTAL	3080

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
17/891,899	08/19/2022	Yuk-Ming Dennis Lo	080015-028430US-1343271

CONFIRMATION NO. 1331

20350
Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

FORMALITIES LETTER

OC000000053746505

Date Mailed: 02/14/2023

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Filing Date Granted

An application number and filing date have been accorded to the above-identified nonprovisional application. The application is informal since it does not comply with the regulations for the reason(s) indicated below. Applicant is given **TWO MONTHS** from the date of this Notice within which to file all required items below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- **Drawing Items**

  The drawings submitted on 08/19/2022 are not acceptable because:
  - Figure(s) 133 contain figure or view numbers that have incorrect orientation. Reference characters, sheet numbers, and view numbers must be oriented in the same direction as the view. See 37 CFR 1.84(p)(1). Replacement drawing sheets in compliance with 37 CFR 1.84 and 1.121(d) having all labeling oriented in the same direction are required.

Replies must be received in the USPTO within the set time period or must include a proper Certificate of Mailing or Transmission under 37 CFR 1.8 with a mailing or transmission date within the set time period. For more information and a suggested format, see Form PTO/SB/92 and MPEP 512.

Replies should be addressed to:

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

REPLY VIA ELECTRONIC FILING

A reply may be submitted via electronic filing only by registered users of the appropriate USPTO patent electronic filing system and must include:

- A separate document having the document description, "**Applicant response to Pre-Exam Formalities Notice**"; and
- the item(s) for which a reply period is set forth above.

For more information about electronic filing, contact the Patent Electronic Business Center (EBC) 1-866-217-9197 or visit https://www.uspto.gov/ebc

NOTE: If the reply is not being submitted electronically, a copy of this notice must be included.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/hchin/

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

U.S. PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code [1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1	8467976	B2	Jun 18, 2013	Lo et al.	
	2	9116118	B2	Aug 25, 2015	Turner et al.	
	3	9175338	B2	Nov 3, 2015	Flusberg et al.	
	4	9175341	B2	Nov 3, 2015	Flusberg et al.	
	5	9175348	B2	Nov 3, 2015	Korlach et al.	
	6	10438691	B2	Oct 8, 2019	Kim et al.	
	7	10465232	B1	Nov 5, 2019	Wu et al.	
	8	10468121	B2	Nov 5, 2019	Kermani et al.	
	9	10590484	B2	Mar 17, 2020	Korlach et al.	

U.S. PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code [1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	10	20070042365	A1	Feb 22, 2007	Millar et al.	
	11	20100221716	A1	Sep 2, 2010	Flusberg et al.	
	12	20110183320	A1	Jul 28, 2011	Flusberg et al.	
	13	20130230909	A1	Sep 5, 2013	Pan et al.	
	14	20140004511	A1	Jan 2, 2014	Korlach et al.	
	15	20160017419	A1	Jan 21, 2016	Chiu et al.	
	16	20170037464	A1	Feb 9, 2017	Turner et al.	
	17	20170159132	A1	Jun 8, 2017	Okino et al.	
	18	20170233802	A1	Aug 17, 2017	Flusberg et al.	
	19	20180201993	A1	Jul 19, 2018	Turner et al.	
	20	20190024162	A1	Jan 24, 2019	Korlach et al.	

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T[5]
	21	2002214811	AU	B2	Sep 9, 2004	Human Genetic Signatures Pty Ltd.		
	22	2005024053	WO	A1	Mar 17, 2005	Human Genetic Signatures Pty Ltd.		
	23	2009013492	WO	A1	Jan 29, 2009	The Chinese University of Hong Kong		
	24	2010027484	WO	A2	Mar 11, 2010	Pacific Biosciences of California, Inc.		
	25	2010068289	WO	A2	Jun 17, 2010	Pacific Biosciences of California, Inc.		
	26	2012170429	WO	A1	Dec 13, 2012	United States Postal Service		
	27	2013170429	WO	A1	Nov 21, 2013	BGI Genomics Co., Ltd.	English Abstract Attached	
	28	2013188846	WO	A1	Dec 19, 2013	Harry Stylli		
	29	104053789	CN	A	Sep 17, 2014	BGI Health Service Co., Ltd.	English Abstract Attached	
	30	2014153757	WO	A1	Oct 2, 2014	BGI Shenzhen	English Abstract Attached	
	31	104053789	CN	B	Feb 10, 2016	BGI Genomics Co., Ltd.	English Abstract Attached	
	32	2012380221	AU	B2	Sep 29, 2016	BGI Genomics Co., Ltd.		
	33	2017012544	WO	A1	Jan 26, 2017	The Chinese University of Hong Kong		
	34	101896167	KR	B1	Sep 7, 2018	The Chinese University of Hong Kong	English Abstract of corresponding document: W02009013492 (A1) Attached	

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	35	Machine Learning, Wikipedia, Available Online at: https://en.wikipedia.org/w/index.php?title=Machine learning&oldid=910906899#Free and open-source software, Assessed on internet at: October 12, 2021, 21 pages.	
	36	U.S. Patent Application No. 16/995,607, Non-Final Office Action mailed on December 30, 2020, 19 pages.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

	37	U.S. Patent Application No. 17/379,544, Non-Final Office Action mailed on February 4, 2022, 10 pages.	
	38	Australian Application No. 2020323958, First Examination Report mailed on May 31, 2021, 3 pages.	
	39	Australian Application No. 2022231719, First Examination Report mailed on September 28, 2022, 4 pages.	
	40	BEAULAURIER et al., Deciphering Bacterial Epigenomes Using Modern Sequencing Technologies, Nature Reviews Genetics, Volume 20 Issue 3, March 2019, pp. 157-172.	
	41	BLOW et al., The Epigenomic Landscape of Prokaryotes, PLOS Genetics, Volume 12, No. 2, February 12, 2016, pp. 1-28.	
	42	CLARK et al., Characterization of DNA Methyltransferase Specificities Using Single-Molecule, Real-Time DNA Sequencing, Nucleic Acids Research, Volume 40, No. 4, Available online at: https://doi.org/10.1093/nar/gkr1146, 2012, 12 pages	
	43	CLARK et al., Enhanced 5-Methylcytosine Detection in Single-Molecule, Real-Time Sequencing via Tet1 Oxidation, BMC Biology, Volume 11, No. 4, 2013, 10 pages.	
	44	EID et al., Real-Time DNA Sequencing from Single Polymerase Molecules, Science, Volume 323, No. 5910, January 2, 2009, pp. 133-138.	
	45	European Application No. 20853612.8, Extended European Search Report Action mailed on October 25, 2021, 15 pages.	
	46	FENG et al., Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic, PLOS Computational Biology, Volume 9, No. 3, Available online at: https://doi.org/10.1371/journal.pcbi.1002935, March 2013, 10 pages.	
	47	FENG et al., qDNAmod: A Statistical Model-based Tool to Reveal Intercellular Heterogeneity of DNA Modification from SMRT Sequencing Data, Nucleic Acids Research, Volume 42, No. 22, November 17, 2014, pp. 13488-13499.	
	48	FLUSBERG et al., Direct Detection of DNA Methylation During Single-Molecule, Real-Time Sequencing, Nature Methods, Volume 7, No. 6, Available Online at: htt12s://12ubmed.ncbi.nlm.nih.gov/20453 866/, June 2010, pp. 461-465,	
	49	United Kingdom Application No. 2102808.9, First Examination Report mailed on April 12, 2021, 6 pages.	
	50	United Kingdom Application No. 2102808.9, Office Action mailed on September 10, 2021, 4 pages.	
	51	United Kingdom Application No. 2102808.9, Office Action mailed on June 23, 2021, 8 pages.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

	52	United Kingdom Application No. 2103003.6, Combined Search and Examination Report mailed on March 29, 2021, 5 pages.	
	53	United Kingdom Application No. 2103010.1, Combined Search and Examination Report mailed on April 15, 2021, 4 pages.	
	54	United Kingdom Application No. 2201811.3, Combined Search and Examination Report mailed on March 1, 2022, 4 pages.	
	55	United Kingdom Application No. 2201821.2, Search Report mailed on February 24, 2022, 4 pages.	
	56	United Kingdom Application No. 2210631.4, Search Report mailed on November 7, 2022, 4 pages.	
	57	United Kingdom Application No. 2210633.0, Search Report mailed on August 26, 2022, 4 pages.	
	58	United Kingdom Application No. 2216178.0, Search Report mailed on November 15, 2022, 4 pages.	
	59	HE et al., TETI Promotes Leukemic Growth of AMLI-ETO+ AML Via 5-Hydroxymethylcytosine Marks and its Oncogenic Role and High Expression can be Antagonized Using Olaparib, an Inhibitor of its Binding Partner PARPI, Experimental Hematology, Volume 53, 2017, pp. S92-S93.	
	60	Japanese Application No. 2021-514525, Office Action mailed on July 26, 2022, 7 pages (4 pages of English Translation and 3 pages of Original Document).	
	61	KILE et al., Correlation of Global and Gene-Specific DNA Methylation in Maternal-Infant Pairs, PLoS One, Volume 5, No. 10, October 29, 2010, 9 pages.	
	62	LIU et al., Bisulfite-Free Direct Detection of 5-Methylcytosine and 5- Hydroxymethylcytosine at Base Resolution, Nature Biotechnology Letters, Volume 37, April 2019, pp. 424-429.	
	63	LIU et al., Detection of DNA Base Modifications by Deep Recurrent Neural Network on Oxford Nanopore Sequencing Data, Nature Communications, Volume 10, No. 2449, June 4, 2019, 11 pages.	
	64	NI et al., Deepsignal: Detecting DNA Methylation State from Nanopore Sequencing Reads Using Deep-Learning, Bioinformatics, Volume 35, No. 22, 2019, pp. 4586-4595.	
	65	International Application No. PCT/CN2020/109602, International Search Report and Written Opinion mailed on November 19, 2020, 11 pages.	
	66	SCHADT et al., Modeling Kinetic Rate Variation in Third Generation DNA Sequencing Data to Detect Putative Modifications to DNA Bases, Genome Research, Volume 23, No. 1, January 2013, pp. 129-141.	
	67	Singapore Application No. 11202101070Q, Written Opinion mailed on October 28, 2021, 9 pages.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

	68	SUGAI et al., Analysis of the DNA Methylation Level of Cancer-related Genes in Colorectal Cancer and the Surrounding Normal Mucosa, Clinical Epigenetics, Volume 9, No. 55, May 18, 2017, pp. 1-10.	
	69	SUMMERER et al., High-Throughput DNA Sequencing Beyond the Four-Letter Code: Epigenetic Modifications Revealed by Single-Molecule Bypass Kinetics, ChemBioChem, Volume 11, No. 8, 2010, pp. 2499-2501.	
	70	SUZUKI et al., Agin: Measuring the Landscape of CpG Methylation of Individual Repetitive Elements, Bioinformatics, Volume 32, No.19, June 17, 2016, pp. 2911-2919.	
	71	SUZUKI, Observing Comprehensive DNA Methylomes via Single Molecule Real-Time Sequencing: Application to Diploid and Centromeric Methylation, Dissertation submitted to The Graduate School of Frontier Sciences, The University of Tokyo, December 13, 2017, 118 pages.	

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document.
[4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/891,899
	Filing Date	August 19, 2022
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	Not Yet Assigned
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	080015-1343271

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David D. Hsu/	Date (YYYY-MM-DD)	2023-02-02
Name/Print	David D. Hsu	Registration Number	71659

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

(12) STANDARD PATENT (11) Application No. **AU 2002214811 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Detection of methylated DNA molecules

(51)[7] International Patent Classification(s)
C12Q 001/68

(21) Application No: **2002214811** (22) Date of Filing: **2001.11.12**

(87) WIPO No: **WO02/38801**

(30) Priority Data

(31) Number	(32) Date	(33) Country
PR 1425	**2000.11.13**	**AU**

(43) Publication Date: **2002.05.21**
(43) Publication Journal Date: **2002.07.25**
(44) Accepted Journal Date: **2004.09.09**

(71) Applicant(s)
Human Genetic Signatures Pty Ltd

(72) Inventor(s)
Molloy, Peter; Grigg, Geoffrey Walter; Millar, Douglas Spencer

(74) Agent / Attorney
Allens Arthur Robinson, The Chifley Tower 2 Chifley Square, Sydney, NSW, 2000

(56) Related Art
AU 68195/00

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/38801 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/AU01/01465

(22) International Filing Date:
12 November 2001 (12.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PR 1425 13 November 2000 (13.11.2000) AU

(71) Applicant *(for all designated States except US)*: **HUMAN GENETIC SIGNATURES PTY LTD** [AU/AU]; 63 Kananook Avenue, Bayview, NSW 2104 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GRIGG, Geoffrey, Walter** [AU/AU]; 352 Burns Bay Road, Linley Point, NSW 2066 (AU). **MOLLOY, Peter** [AU/AU]; 41 Bellevue Street, Chatswood, NSW 2067 (AU). **MILLAR, Douglas, Spencer** [GB/AU]; 8 Rowland Street, Revesby, NSW 2212 (AU).

(74) Agent: **ALLENS ARTHUR ROBINSON**; Patent & Trade Marks Attorneys, Chifley Tower, 2 Chifley Square, Sydney, NSW 2000 (AU).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/38801 A1

(54) Title: DETECTION OF METHYLATED DNA MOLECULES

(57) Abstract: A method for detecting presence of a target DNA in a sample, the method comprising: (a) treating a sample containing DNA with an agent that modifies unmethylated cytosine; (b) providing to the treated sample a detector ligand capable of binding to a target region of DNA and allowing sufficient time for a detector ligand to bind to a target DNA; and (c) measuring binding of the detector ligand to DNA in the sample to determine the presence of the target DNA in a sample.

DETECTION OF NUCLEIC ACID MOLECULES

Technical Field

This invention relates to DNA hybridisation assays and in particular to an improved oligonucleotide or peptide nucleic acid (PNA) assay. The invention also relates to methods for distinguishing specific base sequences including 5-methyl cytosine bases in DNA using these assays.

Background Art

A number of procedures were available for the detection of specific nucleic acid molecules. These procedures typically depend on sequence-dependent hybridisation between the target DNA and nucleic acid probes which may range in length from short oligonucleotides (20 bases or less) to sequences of many kilobases.

For direct detection, the target DNA is most commonly separated on the basis of size by gel electrophoresis and transferred to a solid support prior to hybridisation with a probe complementary to the target sequence (Southern and Northern blotting). The probe may be a natural nucleic acid or analogue such as PNA or locked nucleic acid (LNA). The probe may be directly labelled (eg. with ^{32}P) or an indirect detection procedure may be used. Indirect procedures usually rely on incorporation into the probe of a "tag" such as biotin or digoxigenin and the probe is then detected by means such as enzyme-linked substrate conversion or chemiluminescence.

Another method for direct detection of nucleic acid that has been used widely is "sandwich" hybridisation. In this method, a capture probe is coupled to a solid support and the target DNA, in solution, is hybridised with the bound probe. Unbound target DNA is washed away and the bound DNA is detected using a second probe that hybridises to the target sequences. Detection may use direct or indirect methods as outlined above. The "branched DNA" signal detection system is an example that uses the sandwich hybridization principle (Urdea Ms Branched DNA signal amplification. Biotechnology 12: 926-928).

A rapidly growing area that uses nucleic acid hybridisation for direct detection of nucleic acid sequences is that of DNA micro-arrays (Young RA Biomedical discovery with DNA arrays. Cell 102: 9-15 (2000); Watson, A New tools. A new breed of high tech detectives. Science 289:850-854 (2000)). In this process, individual nucleic acid species, that may range from oligonucleotides to longer sequences such as cDNA clones, were fixed to a solid support in a grid pattern. A tagged or labelled nucleic acid population is then hybridised with the array and the level of hybridisation with each spot in the array is quantified. Most commonly, radioactively or fluorescently-labelled nucleic acids (eg. cDNAs) were used for hybridisation, though other detection systems were employed.

The most widely used method for amplification of specific sequences from within a population of nucleic acid sequences is that of polymerase chain reaction (PCR) (Dieffenbach C and Dveksler G eds. PCR Primer: A Laboratory Manual. Cold Spring Harbor Press, Plainview NY). In this amplification method, oligonucleotides, generally 15 to 30 nucleotides in length on complementary strands and at either end of the region to be amplified, were used to prime DNA synthesis on denatured single-stranded DNA. Successive cycles of denaturation, primer hybridisation and DNA strand synthesis using thermostable DNA polymerases allows exponential amplification of the sequences between the primers. RNA sequences can be amplified by first copying using reverse transcriptase to produce a cDNA copy. Amplified DNA fragments can be detected by a variety of means including gel electrophoresis, hybridisation with labelled probes, use of tagged primers that allow subsequent identification (eg. by an enzyme linked assay), use of fluorescently-tagged primers that give rise to a signal upon hybridisation with the target DNA (eg. Beacon and TaqMan systems).

As well as PCR, a variety of other techniques have been developed for detection and amplification of specific sequences. One example is the ligase chain reaction (Barany F Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc. Natl. Acad. Sci. USA 88:189-193 (1991)).

Currently the method of choice to detect methylation changes in DNA, such as were found in the GSTP1 gene promoter in prostate cancer, were dependent on PCR amplification of such sequences after bisulfite modification of

DNA. In bisulfite-treated DNA, cytosines were converted to uracils (and hence amplified as thymines during PCR) while methylated cytosines were non-reactive and remain as cytosines (Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL and Paul CL. A genomic sequencing protocol which yields a positive display of 5-methyl cytosine residues in individual DNA strands. PNAS 89: 1827-1831 (1992); Clark SJ, Harrison J, Paul CL and Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 22: 2990-2997 (1994)). Thus (after bisulfite treatment) DNA containing 5-methyl cytosine bases will be different in sequence from the corresponding unmethylated DNA. This report results in the basis of the bisulfite method for sequencing 5-methyl cytosine residues in DNA. Surprisingly, several years later this assay was used as the basis of a PCR assay for the methylation status of CpG islands in US 5786146. Primers may be chosen to amplify non-selectively a region of the genome of interest to determine its methylation status, or may be designed to selectively amplify sequences in which particular cytosines were methylated (Herman JG, Graff JR, Myohanen S, Nelkin BD and Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. PNAS 93:9821-9826 (1996)).

Alternative methods for detection of cytosine methylation include digestion with restriction enzymes whose cutting is blocked by site-specific DNA methylation, followed by Southern blotting and hybridisation probing for the region of interest. This approach is limited to circumstances where a significant proportion (generally >10%) of the DNA is methylated at the site and where there is sufficient DNA, usually 10 μg, to allow for detection. Digestion with restriction enzymes whose cutting is blocked by site-specific DNA methylation, followed by PCR amplification using primers that flank the restriction enzyme site(s). This method can utilise smaller amounts of DNA but any lack of complete enzyme digestion for reasons other than DNA methylation can lead to false positive signals.

Recently, peptide nucleic acids (PNA) in which the entire deoxyribose-phosphate backbone has been exchanged with a structurally homomorphous uncharged polyamide backbone composed of N-(2-aminoethyl)glycine units have

been developed (Ray A and Norden B. Peptide nucleic acid (PNA): its medical and biotechnical applications and for the future. FASEB J 14: 1041-1060 (2000)). These compounds contain the same number of backbone bonds between the bases (i.e. 6) and the same number of bonds from the backbone to the base (i.e. 3) as in DNA. PNA oligomers have been found to bind with high affinity and sequence specificity to both complementary RNA and DNA and a number of oligonucleotide-dependent enzymatic functions have been inhibited on forming PNA/DNA or PNA/RNA complexes. PNAs demonstrate a higher binding affinity than their equivalent oligonucleotides and mismatches of PNAs with complementary nucleotide sequences cause a more profound lowering of melting temperature than is seen with oligonucleotides. PNAs have also shown a number of special properties, one of which is that homopyrimidine PNAs bind to double-stranded DNA with displaced strand analogous to a D-loop. More recently, Neilsen (Nielsen PE. Peptide nucleic acids as therapeutic agents. Curr. Open Struct. Biol. 9: 353-357 (1999) has reported that a homopurine PNA binds to double-stranded DNA with displacement of the non-complementary strand, resulting in formation of a PNA/DNA duplex and a displaced D-loop. However, unlike homopyrimidine PNAs, the homopurine PNA/DNA duplex is not then further stabilised by triplex formation. Hence, PNA offers both antisense and antigene strategies for regulating gene expression.

The present inventors have now developed methods utilizing ligands for the sensitive and specific detection of DNA which do not require PCR amplification.

Disclosure of Invention

In a first aspect, the present invention provides a method for detecting presence of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing to the treated sample a detector ligand capable of binding to a target region of DNA and allowing sufficient time for a detector ligand to bind to a target DNA; and

(c) measuring binding of the detector ligand to DNA in the sample to determine the presence of the target DNA in a sample.

In a second aspect, the present invention provides a method for estimating extent of methylation of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing to the treated sample a detector ligand capable of distinguishing between methylated and unmethylated cytosine of DNA and allowing sufficient time for a detector ligand to bind to a target DNA; and

(c) detecting binding of the detector ligand to DNA in the sample such that the degree or amount of binding is indicative of the extent of methylation of the target DNA.

In step (b), two detector ligands can be used where one ligand is capable of binding to a region of DNA that contains one or more methylated cytosines and the other ligand capable of binding to a corresponding region of DNA that contains no methylated cytosines. As a sample can contain many copies of a target DNA, often the copies have different amounts of methylation. Accordingly, the ratio of binding of the two ligands will be proportional to the degree of methylation of that DNA target in the sample. The two ligands can be added together in the one test or can be added in separate duplicate tests. Each ligand can contain a unique marker which can be detected concurrently or separately in the one test or have the same marker and detected individually in separate tests.

In a third aspect, the invention provides a method for detecting the presence of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing a support to which is bound a capture ligand which is capable of recognising a first part of a target DNA sequence;

(c) contacting the support with the treated sample for sufficient time to allow DNA to bind to a capture ligand such that target DNA in the sample binds to the support via the capture ligand;

(d) contacting the support with a detector ligand capable of recognising a second part of the target DNA sequence and allowing sufficient time for a detector ligand to bind to a target DNA bound to a support; and

(e) measuring binding of the detector ligand to DNA bound to the support to determine the presence of the target DNA in the sample.

In a fourth aspect, the present invention provides a method for estimating extent of methylation of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing a support to which is bound a capture ligand which is capable of recognising a first part of a target DNA sequence;

(c) contacting the support with the treated sample for sufficient time to allow DNA to bind to a capture ligand such that target DNA in the sample binds to the support via the capture ligand;

(d) contacting the support with a detector ligand capable of distinguishing between methylated and unmethylated cytosine of DNA such that the detector ligand binds to any target DNA on the support; and

(e) detecting binding of the detector ligand to the support such that the degree or amount of binding is indicative of the extent of methylation of the target DNA.

Preferably, the capture ligand is selected from peptide nucleic acid (PNA) probe, oligonucleotide, modified oligonucleotide, single stranded DNA, RNA, aptamer, antibody, protein, peptide, a combination thereof, or chimeric versions thereof.

More preferably, the capture ligand is a PNA probe or an oligonucleotide probe. Even more preferably, the capture ligand is a PNA probe.

The support can be any suitable support such as a plastic materials, fluorescent beads, magnetic beads, synthetic or natural membranes, latex beads, polystyrene, column supports, glass beads or slides, nanotubes, fibres or other organic or inorganic supports. Preferably, the support is a magnetic bead or a fluorescent bead.

The solid substrate is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs or microplates, or any other surface suitable for conducting an assay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing the molecule to the insoluble carrier.

In a preferred form, step (b) comprises a plurality capture ligands arrayed on a solid support. The array may contain multiple copies of the same ligand so as to capture the same target DNA on the array or may contain a plurality of different ligands targeted to different DNA so as to capture a plurality of target DNA molecules on the array. Typically, the array contains from about 10 to 10,000 capture ligands. In one form, the array has less than about 500 capture ligands. It will be appreciated, however, that the array can have any number of capture ligands.

In one form, capture oligonucleotide probes or capture PNA probes can be placed on an array and used to capture bisulfite-treated DNA to measure methylated states of DNA. Array technology is well known and has been used to detect the presence of genes or nucleotide sequences in untreated samples. The present invention, however, can extend the usefulness of array technology to provide valuable information on methylation states of many different sources of DNA.

The sample can be any biological sample such as blood, urine. faeces, semen, cerebrospinal fluid, cells or tissue such as brain, colon, urogenital, lung, renal, hematopoietic, breast, thymus, testis, ovary, or uterus, environmental samples, microorganisms including bacteria, virus, fungi, protozoan, viroid and the like.

In preferred forms, the sample is blood, colorectal tissue, brain or prostate tissue.

Preferably, the modifying agent is capable of modifying unmethylated cytosine but not methylated cytosine. The agent is preferably is selected from bisulfite, acetate and citrate. Preferably, the agent is sodium bisulfite and cytosine is modified to uracil.

The term "modifies" as used herein means the conversion of an unmethylated cytosine to another nucleotide which will distinguish the unmethylated from the methylated cytosine. Preferably, the agent modifies unmethylated cytosine to uracil. Preferably, the agent used for modifying unmethylated cytosine is sodium bisulfite, however, other agents that similarly modify unmethylated cytosine, but not methylated cytosine can also be used in the method of the invention. Sodium bisulfite ($NaHSO_3$) reacts readily with the 5,6-double bond of cytosine, but poorly with methylated cytosine. Cytosine reacts with the bisulfite ion to form a sulfonated cytosine reaction intermediate which is susceptible to deamination, giving rise to a sulfonated uracil. The sulfonate group can be removed under alkaline conditions, resulting in the formation of uracil. Thus all unmethylated cytosines will be converted to uracil while methylated cytosines will be protected from conversion so that ligands can be prepared that will recognise sequences containing cytosine or corresponding sequences containing uracil. The ratio of binding of the two probes can provide an accurate measure of the degree of methylation in a given DNA. Importantly, there is no need to amplify the DNA to obtain the required information thus overcoming potential errors and resulting in a faster and more simple assay amenable to automation.

In a preferred form, the detector ligand is directed to a CpG- or CNG-containing region of DNA, where N designates any one of the four possible bases

A, T, C, or G. Preferably, the CpG- or CNG- containing region of DNA is in a regulatory region of a gene or an enhancer of any regulatory element or region. This region includes promoter, enhancer, oncogene, or other regulatory element which activity is altered by environmental factors including chemicals, toxins, drugs, radiation, synthetic or natural compounds and microorganisms or other infectious agents such as viruses, bacteria, fungi and prions. For example, the promoter or regulatory element can be a tumour suppressor gene promoter, oncogene or any other element that may control or influence one or more genes implicated in a disease state or changing normal state such as aging.

The presence of methylated CpG- or CNG- containing region of DNA in a specimen can be indicative of a cell proliferative disorder. The disorder can include low grade astrocytoma, anaplastic astrocytoma, glioblastoma, medulloblastoma, colon cancer, lung cancer, renal cancer, leukemia, breast cancer, prostate cancer, endometrial cancer and neuroblastoma.

In order to assist in the reaction of the DNA modifying agent optional additives such as urea, methoxyamine and mixtures thereof can be added.

Step (b) is typically used to capture a DNA of interest which will be analysed for methylation in subsequent steps of the method. Often a sample will contain genomic DNA from a cell source and that only one or a few genes will be of interest. Thus, step (b) allows the capture and concentration of DNA of interest. Preferably a first PNA or oligonucleotide probe is used in step (b).

In one preferred form, step (b) comprises a plurality of capture ligands arrayed on a solid support. The array may contain multiple copies of the same ligand so as to capture the same target DNA on the array for subsequent testing. Alternatively, the array may contain a plurality of different capture ligands targeted to different DNA molecules so as to capture many different target DNA samples on the array for subsequent testing. In one preferred form, the capture ligands are oligonucleotides or PNA molecules.

In step (d), two detector ligands can be used where one ligand is capable of binding to a region of DNA that contains one or more methylated cytosines and the second ligand is capable of binding to a corresponding region of DNA that contains no methylated cytosines. A sample can contain many copies of a target

DNA with the copies having different amounts of methylation. Accordingly, the ratio of binding of the two ligands will be proportional to the degree of methylation of that DNA target in the sample. The two ligands can be added together in the one test or can be added in separate duplicate tests. Each ligand can have an unique marker which can be detected concurrently or separately in the one test or have the same marker and detected individually in separate tests.

In order to detect binding of the detector ligand to a target DNA, preferably the ligand has a detectable label attached thereto. The presence of bound label being indicative of the extent of binding of the ligand. Suitable labels include fluorescence, radioactivity, enzyme, hapten and dendrimer.

The detector ligands used in the invention for detecting CpG- or CNG-containing DNA in a sample, after bisulfite modification, can specifically distinguish between untreated DNA, methylated, and unmethylated DNA. Detector ligands in the form of oligonucleotide or PNA probes for the non-methylated DNA preferably have a T or A in the 3' CG or CNG pair to distinguish it from the C retained in methylated DNA.

The probes of the invention were designed to be "substantially" complementary to one strand of the genomic locus to be tested and include the appropriate G or C nucleotides. This means that the primers must be sufficiently complementary to hybridize with a respective region of interest under conditions which allow binding. In other words, the probes should have sufficient complementarity with the 5' and 3' flanking sequences to hybridize therewith.

The PNA probes of the invention may be prepared using any suitable method known to the art. Typically, the PNA probes were prepared according to methods outlined in US 6110676 (Coull et al 2000), incorporated herein by reference

The methods according to the present invention relating to methylation states of target DNA can use any DNA sample, in purified or unpurified form, as the starting material, provided it contains, or is suspected of containing, the specific DNA sequence containing the target region (usually CpG or CNG). Typically, unamplified samples are used in the methods according to the present invention.

The DNA-containing specimen used for detection of methylated CpG or CNG may be from any source and may be extracted by a variety of techniques such as that described by Maniatis, et al (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp 280, 281, 1982).

Where the DNA in the sample contains two strands, it is necessary to separate the strands of the DNA before it can be modified. Strand separation can be effected either as a separate step or simultaneously with chemical treatment. This strand separation can be accomplished using various suitable denaturing conditions, including physical, chemical, or enzymatic means, the word "denaturing" includes all such means. One physical method of separating DNA strands involves heating the DNA until it is denatured. Typical heat denaturation may involve temperatures ranging from about 80° to 105°C for times ranging from about 1 to 10 minutes. Strand separation may also be induced by an enzyme from the class of enzymes known as helicases or by the enzyme RecA, which has helicase activity, and in the presence of riboATP, is known to denature DNA. The reaction conditions suitable for strand separation of DNA with helicases were described by Kuhn Hoffmann-Berling (CSH-Quantitative Biology, 43:63, 1978) and techniques for using RecA were reviewed in C. Radding (Ann. Rev. Genetics, 16:405-437, 1982.

The detectable label may be fluorescent, or radioactive or contain a second label or marker in the form of a microsphere. The fluorescent or radioactive microsphere may be covalently bound to the capture or detector ligand.

In the case of a capture PNA ligand, the DNA binding can be detected via the phosphate groups thereby ensuring highly specific binding to the DNA and not to the negatively charged ligand or uncharged PNA.

The reagent is preferably a cationic molecule which binds to the DNA electrostatically. The detectable label attached thereto may be a fluorescent or radioactive molecule.

Preferably the specificity of hybridization to target DNA is used to discriminate between methylated cytosines and unmethylated cytosines.

The present invention makes particular use of the fact that PNA molecules have no net electrical charge while DNA, because of its phosphate backbone, are highly negatively charged. Detection of bound PNA probes can be a simple molecule such as a positively charged fluorochrome, multiple molecules of which will bind specifically to the DNA in proportion to its length and can be directly detected. Many suitable fluorochromes that bind to DNA, some selective for single-stranded DNA, and that differ in their excitation and emission wavelengths were known. The detection system could also be an enzyme carrying a positively charged region that will selectively bind to the DNA and that can be detected using an enzymatic assay, or a positively charged radioactive molecule that binds selectively to the captured DNA.

Using PNA probes as one of the ligands in this procedure has very significant advantages over the use of oligonucleotide probes. PNA binding reaches equilibrium faster and exhibits greater sequence specificity and, as PNAs are uncharged, they bind the target DNA molecules with a higher binding coefficient.

As the invention can use direct detection methods, they give a true and accurate measure of the amount of a target DNA in a sample. The methods were not confounded by potential bias inherent in methods that rely for signal amplification on processes such as PCR, where the enzymes commonly used in such procedures can introduce systematic bias through differential rates of amplification of different sequences.

In a fifth aspect, the present invention provides a method for detecting a methylated CpG- or CNG-containing DNA, the method comprising:

A method for detecting a methylated CpG- or CNG-containing DNA, the method comprising:

(a) treating a sample containing DNA with bisulfite to modify unmethylated cytosine to uracil in the DNA;

(b) providing to the treated sample a detector PNA ligand capable of distinguishing between methylated and unmethylated cytosine of DNA; and

(c) detecting the methylated DNA based on the presence or absence of binding of the detector PNA ligand.

In one preferred from, the method comprises:

(a) treating a DNA-containing specimen with bisulfite to modify unmethylated cytosine to uracil,

(b) providing to the treated sample a detector ligand capable of binding to a methylated CpG- or CNG-containing DNA but not to a corresponding unmethylated CpG- or CNG-containing DNA; and

(c) detecting binding of the ligand to DNA in the sample such that binding is indicative of methylation of the DNA.

Preferably, the detector ligand is a peptide nucleic acid (PNA) probe.

In a preferred from, the invention provides a method for estimating extent of methylation of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with bisulfite to modify unmethylated cytosine to uracil;

(b) providing a solid support in the form of a magnetic bead to which is bound a capture PNA or oligonucleotide ligand which is capable of recognising a first part of a target DNA sequence;

(c) contacting the support with the treated sample suspected of containing the target DNA such that target DNA in the sample binds to the support via the capture ligand;

(d) contacting the support with a detector PNA ligand capable of distinguishing between methylated and unmethylated cytosine of DNA; and

(e) determining the extent of methylation of the DNA bound to the support by measuring the amount of bound detector ligand.

In a sixth aspect, the present invention relates to use of an agent that modifies unmethylated cytosine but not methylated cytosine and one or more ligands, preferably one or more peptide nucleic acid (PNA) probes, capable of distinguishing between methylated and unmethylated cytosine of DNA in methods for assaying methylation of target DNA.

There as a number of detector systems and instruments available for detecting or measuring fluorescence or radioactivity. Improvements and advancement in instrumentation is being made by a number of manufacturers. It will be appreciated that many different measuring instruments can be used for the present invention. For example, Multi Photon Detection is a proprietary system for the detection of ultra low amounts of selected radioisotopes. It is 1000 fold more sensitive than existing methods. It has a sensitivity of 1000 atoms of iodine 125, with quantitation of zeptomole amounts of biomaterials. It requires less than 1 picoCurie of isotope which is 100 times less activity than in a glass of water. A family of MPD instruments already exists for measuring radioactivity in a sample. They consist of instruments that are configured for 96 well, 384 well and higher. MPD uses coincident multichannel detection of photons coupled with computer controlled electronics to selectively count only those photons that are compatible with an operator-selected radioisotope. As many different isotopes can be used, this is a multicolor system. The MPD imager system is at least 100 fold more sensitive than a phosphor imager. Such instrumentation would be particularly suitable in the detection part of the present invention where ligands or supports are made radioactive.

Beads containing capture or detector ligands bound thereto can be processed or measured by cell sorters which measure fluorescence. Examples or suitable instruments include flow cytometers and modified versions thereof.

The methods according to the present invention are particularly suitable for scaling up and automation for processing many samples.

Notwithstanding the above, the methods described can be used in conjunction with such amplification procedures if it is necessary to amplify limiting amounts of DNA in order provide enough material for detection.

Methylated DNA: In a particular adaptation as detailed in the present invention, the methods can be used to distinguish the presence of methylated cytosines in DNA that has been treated with sodium bisulfite. As cytosines were converted to uracils while methyl cytosines remain unreacted, the sequence of bisulfite-treated DNA derived from methylated and unmethylated molecules is different. By choosing a specific PNA ligand (5 to 40 residues long, preferably 15 $\pm$ 5 residues

long) to selected target regions the specificity of hybridisation can be used to discriminate between methylated cytosines at CpG or CNG sites (which remain as cytosines) and unmethylated CpG or CNG sites where the cytosine is converted to uracil, while ensuring that only molecules in which cytosines that were not in CpG or CNG sites have fully reacted and been converted to uracils were assessed.

Methylated cytosines at other sites can similarly be detected. Appropriate PNA probes can be used as controls to identify the presence of molecules that have not reacted completely with bisulfite (one or more cytosines not converted to uracil). It will be appreciated, however, that other ligands which can differentiate between the methylation states of DNA can be used in a similar manner.

The methods were amenable for use in a variety of formats including multiwell plates, micro-arrays and particles in suspension. The appropriate selection of specific ligands for use in an array format can allow for the simultaneous determination of the methylation state of individual cytosines in multiple target regions.

Polymorphism/mutation detection: The methods according to the present invention can be applied to the discrimination of mutant alleles of a gene where the sequence of the capture ligand and/or the detector ligand will match with one allele but mismatch with the other.

DNA Quantification: By using the methods according to the present invention, it is possible to directly determine within a DNA population the proportion of molecules having one sequence versus another at a particular region. This can be done by coupling ligands representing the alternate forms of the sequence to supports such as microspheres charged with different fluorochromes or radioactive molecules. Such differences in sequence may be differences in the original base sequence of the gene or differences in base sequence in bisulfite-treated DNA that were due to differences in methylation in the original DNA.

Cell quantification: The methods can be applied to determining the ratio of cells in a population (such as in cancer and normal cells) that differ in base sequence at a particular site in the genome.

Variations: The methods were amenable for use in a variety of formats including multiwell plates, micro-arrays and particles in suspension. The appropriate selection of specific PNA probes for use in an array format can allow for the simultaneous determination of the presence of different DNA sequences, eg. for the determination of the methylation state of individual cytosines in multiple target regions.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed in Australia before the priority date of each claim of this application.

In order that the present invention may be more clearly understood, preferred forms will be described with reference to the following drawings and examples.

Brief Description of the Drawings

Figure 1 shows a general overview of sandwich signal amplification methodology using PNA probes for detection of methylated DNA.

Figure 2 hows a general overview of sandwich signal amplification methodology using PNA probes and magnetic beads for detection of methylated DNA.

Figure 3 shows part of the nucleic acid sequence of the GSTP1 gene and ten PNA probes useful for detecting various methylation states of that gene region.

Figure 4 shows a comparison of the effect of microsphere bead size on hybridisation signal.

Figure 5 shows detection capabilities for prostate cancer cell line and tissue DNA extracts using PNA technology and methods of the invention.

Figure 6 shows effect of PNA concentration on sensitivity of method using ligands bound to micotitre well plates.

Figure 7 shows results of single methylation using Oligreen detection agent.

Figure 8 shows results of detection of methylated DNA sequences in a background of unmethylated sequences.

Figure 9 shows results of detection of unmethylated DNA sequences in a background of methylated sequences.

Figure 10 shows an example of the methylation pattern of GSTP1 in prostate cancer.

Mode(s) for Carrying Out the Invention

Methylation

The amount or degree of methylation of genomic DNA has implications in many conditions such as aging, genetic abnormalities, cancer and other disease states. A number of important implications of methylation states were set out below.

The fusion of Embryonic Stem Cells with adult thymocytes to examine the reprogramming that occurs at the level of DNA methylation after the fusion has been made. The inactive somatic X becomes activated as visualized by whole chromosome examination (Tada *et al.*, 2001; Current Biology, 11, 1553-1558).

Examination of methylation patterns in specific DNA regions in the clinicopathological features of sporadic colorectal cancers, as an inexpensive and accurate way of identifying such tumors (Ward *et al.*, 2001; Gut, 48,821-829), and the methylation patterns in stem cells in human colon crypts (Ro et al., 2001,

Proc Natl Acad Sci, USA, 98, 10519-10521; Yatabe et al., 2001, Proc. Natl. Acad Sci USA, on line edition).

Methylation patterns in prostate cancer, and in cell lines treated with 5-azacytidine in order to reactivate specific genes (Chetcuti *et al.*, 2001, Cancer Research, 61,6331-6334).

Methylation patterns in the various Estrogen receptors in uterus endometrial cancers where gene inactivation via methylation occurs in many cancers but is not at a high frequency in normal individuals (Sasaki et al., 2001, Cancer Research, 61, 3262-3266).

Methylation patterns in bladder cancer (Markl et al., 2001, Cancer Research, 61, 5875-5884).

Methylation patterns in breast cancer (Nielsen et al., 2001, Cancer Letters, 163, 59-69).

Methylation patterns in specific promoters involved in lung and breast cancers (Burbee et al., 2001, J Natl Cancer Institute, 93, 691-699).

Methylation patterns in free DNA in the plasma of patients with esophageal adenocarcinomas (Kawakami et al., 2000, J Natl Cancer Institute, 92, 1805-1811).

Methylation of the CDH1 promoter in hereditary diffuse gastric cancer (Grady et al., 2000, Nature Genetics, 26, 16-17).

Genomic imprinting, in which, for example, a paternal allele of a gene is active, and the maternal allele is inactive, or vice versa. This inactivation is accomplished via methylation changes in the genes involved, or in sequences nearby to them. In essence, DNA regions become methylated in the germ line of one sex, but not in that of another (Mann, 2001, Stem Cells, 19, 287-294).

Genome-wide methylation patterns in studies of cloning of various species (sheep, cattle, goats, pigs and mice), via nuclear transfer or *in vitro* fertilization. Thus the methylation patterns of donor nuclei that were inserted into oocytes vary greatly, and this is thought to be the reason why there is such a high failure rate in current cloning experiments. These differentiated nuclei probably require more reprogramming that less differentiated ones such as in Embryonic Stem Cells

(Kang et al, 2001; Nature Genetics, 28, 173-177; Humphreys et al., 2001, Science, 293, 95-97).

Excessive hyper-methylation patterns in 24 cancer cell lines versus normal tissues (Smiraglia et al., 2001, Human Molecular Genetics, 10, 1413-1419).

Insertion of methylated DNA into a non methylated mini gene construct to examine the effects on gene expression and imprinting (Holmgren et al., 2001, Current Biology, 11, 1128-1130).

Methylation patterns in mature B cell lymphomas, where specific genes were inactivated by methylation (Malone et al., 2001, Proc Natl Acad Sci USA 98, 10404-10409).

Methylation patterns of particular genes in acute myeloid leukemia (Melki et al., 1999, Leukemia, 13, 877-883).

Analysis of the Mecp2 gene in knockout mice. This protein is involved in binding to methylated sites in DNA and is thought to be involved in Rett syndrome, which is an inherited neurological disorder (Guy et al., Nature Genetics, 27, 322-326).

Methylation patterns of 5 specific genes during the normal aging process, and in ulcerative colitis (Issa et al., 2001, Cancer Research, 61, 3573-3577).

Loss of methylation in the processes of apoptosis, which impinge upon signal transduction pathways, cell cycle control, movement of mobile elements within the genome (Jackson-Grusby et al., 2001, Nature Genetics, 27, 31-39).

Comparison of the methylation patterns of promoter and gene regions in different species, such as human and mouse, to determine the evolutionary conservation or lack thereof of CpG islands involved in gene regulation (Cuadrado et al., 2001, EMBO Reports, 21, 586-592).

DNA methylation patterns in testicular sperm at different developmental stages (Manning et al., 2001, Urol Int, 67, 151-155).

Immuno histochemical staining using a monoclonal antibody to analyze DNA methylation patterns (Piyathilake et al., 2000, Biotechnic and Histochem, 75, 251-258).

Differences between the methylation patterns of genes and pseudogenes (Grunau et al., 2000, Human Mol Genet, 9, 2651-2663).

5-methylycytosine content of model invertebrates such as Drosophila melanogaster (Gowher et al., 2000, EMBO J, 19, 6918-6923).

Large scale mapping of human promoters using the methylation patterns of CpG islands (Ioshikhes et al, 2000, Nature Genetics, 26, 61-63).

Induced changes in the processes of chromatin remodelling, DNA methylation and gene expression during mammalian development due to changes in the expression of the ATRX gene which give rise to mental retardation, facial dysmorphism, urogenital abnormalities and alpha thalassemia (Gibbons et al., 2000, Nature Genetics, 24, 368-371).

Boundaries between methylated and unmethylated domains in the promoter region of the GSTP1 gene involved in prostate cancer (Millar et al., 2000, J Biological Chemistry, 275, 24893-24899; Millar et al., 1999, Oncogene, 18, 1313-1324).

Methylation changes during the normal processes of aging (Toyota et al., 1999, Seminars in Cancer Biology, 9, 349-357).

Methylation changes in aging and in atherosclerosis in the cardiovascular system, (Post et al., 1999, Cardiovascular Research, 43, 985-991) and during normal aging and cancers in colorectal mucosa (Ahuja et al., 1998, Cancer Research, 58, 5489-5494).

Methylation patterns in germ cells and sertoli cells in testis (Coffigny et al., 1999, Cytogenet Cell Genets, 87, 175-181).

DNA methylation changes during the development of model vertebrates such as the zebrafish (Macleod et al., 1999, Nature Genetics, 23, 139-140).

Methylation patterns in the promoter regions of the human histo-blood ABO genes (Kominato et al., 1999, J Biol Chem, 274, 37240-37250).

Methylation patterns during mammalian preimplantation development using monoclonal antibodies (Rougier et al., 1999, Genes and Development, 12, 2108-2113).

Methylation patterns induced by various cancer chemotherapeutic drugs (Nyce, 1997, Mutation Research, 386, 153-161; Nyce 1989, Cancer Research, 49, 5829-5836) and the changes in DNA methylation in phenobarbital-induced and spontaneous liver tumors (Ray et al., 1994, Molecular Carcinogenesis 9, 155-166).

Analysis of 5-methycytosine residues in DNA by the bisulfite sequencing method (Grigg, 1996, DNA Sequence, 6, 189-198).

Isolation of CpG islands using a methylated DNA binding column (Cross et al., 1994, Nature Genetics, 6, 236-244).

Is KSHV lytic growth induced by a methylation-sensitive switch? (Laman and Boshoff, Trends Microbiol 2001 Oct; 9(10):464-6). Both latent and lytic growth of Kaposi's sarcoma-associated herpesvirus (KSHV or HHV-8) contribute to its pathogenesis.

As can be seen from the large number of examples of different methylation states and implications provided above, it will be appreciated that the present invention offers a powerful tool for the study of methylation and thus is useful for many aspects of disease and health.

Table 1 shows some examples of solid supports useful for attaching capture ligands of the present invention. Table 2 shows possible choices of detector systems for use in the present invention.

Table 1. Solid supports for attachment of capture ligands

label	**fluoro bead**	**column**	**magnetic bead**	**latex bead**	**p/styrene bead**	**membrane**	**glass**
PNA	+	+	+	+	+	+	+
Oligo	+	+	+	+	+	+	+
RNA	+	+	+	+	+	+	+
Hybrid	+	+	+	+	+	+	+

Table 2. Detection systems for detection ligands

Label	Fluoro Bead	Magnetic bead	latex bead	p/styrene bead	Glass	Aptamer
Pre-label	+					
Florescence	+	+	+	+	+	+
radiolabel	+	+	+	+	+	+
Dendrimer	+	+	+	+	+	+

MATERIALS AND METHODS

Peptide Nucleic Acids (PNAs)

Peptide Nucleic Acids (PNAs) are non-naturally occurring polyamides which can hybridize to nucleic acids (DNA and RNA) with sequence specificity. (See U.S. Pat. No. 5,539,082 and Egholm et al., Nature (1993) 365, 566-568). PNA's are candidates as alternatives/substitutes to nucleic acid probes in probe-based hybridization assays because they exhibit several desirable properties. PNA's are achiral polymers which hybridize to nucleic acids to form hybrids which are more thermodynamically stable than a corresponding nucleic acid/nucleic acid complex (See: Egholm et. al., Nature (1993) 365, 566-568). Being non-naturally occurring molecules, they are not known to be substrates for the enzymes which are known to degrade peptides or nucleic acids. Therefore, PNA's should be stable in biological samples, as well as, have a long shelf-life. Unlike nucleic acid hybridization which is very dependent on ionic strength, the hybridization of a PNA with a nucleic acid is fairly independent of ionic strength and is favoured at low ionic strength under conditions which strongly disfavour the hybridization of nucleic acid to nucleic acid (See: Egholm et. al., Nature, p. 567). The effect of ionic strength on the stability and conformation of PNA complexes has been extensively investigated (See: Tomac et al. J. Am. Chem. Soc. (1996) 118, 5544-5552). Sequence discrimination is more efficient for PNA recognizing DNA than for DNA recognizing DNA (See: Egholm et al., Nature, p.

566). However, the advantages in point mutation discrimination with PNA probes, as compared with DNA probes, in a hybridization assay appears to be somewhat sequence dependent (See: Nielsen et al. Anti-Cancer Drug Design (1993) 8, 53-65). As an additional advantage, PNA's hybridize to nucleic acid in both a parallel and antiparallel orientation, though the antiparallel orientation is preferred (See: Egholm et al., Nature, p. 566).

PNAs are synthesized by adaptation of standard peptide synthesis procedures in a format which is now commercially available. (For a general review of the preparation of PNA monomers and oligomers please see: Dueholm et al., New J. Chem. (1997), 21, 19-31 or Hyrup et. al., Bioorganic & Med. Chem. (1996) 4, 5-23). Labelled and unlabelled PNA oligomers can be purchased (See: PerSeptive Biosystems Promotional Literature: BioConcepts, Publication No. NL612, Practical PNA, Review and Practical PNA, Vol. 1, Iss. 2) or prepared using the commercially available products.

There are indeed many differences between PNA probes and standard nucleic acid probes. These differences can be conveniently broken down into biological, structural, and physico-chemical differences. As discussed above and below, these biological, structural, and physico-chemical differences may lead to unpredictable results when attempting to use PNA probes in applications were nucleic acids have typically been employed. This non-equivalency of differing compositions is often observed in the chemical arts.

With regard to biological differences, nucleic acids are biological materials that play a central role in the life of living species as agents of genetic transmission and expression. Their *in vivo* properties are fairly well understood. PNA, however, is a recently developed totally artificial molecule, conceived in the minds of chemists and made using synthetic organic chemistry. It has no known biological function.

Structurally, PNA also differs dramatically from nucleic acid. Although both can employ common nucleobases (A, C, G, T, and U), the backbones of these molecules are structurally diverse. The backbones of RNA and DNA are composed of repeating phosphodiester ribose and 2-deoxyribose units. In contrast, the backbones of PNA are composed on N-(2-aminoethyl)glycine units.

Additionally, in PNA the nucleobases are connected to the backbone by an additional methylene carbonyl unit.

Despite its name, PNA is not an acid and contains no charged acidic groups such as those present in DNA and RNA. Because they lack formal charge, PNAs are generally more hydrophobic than their equivalent nucleic acid molecules. The hydrophobic character of PNA allows for the possibility of non-specific (hydrophobic/hydrophobic interactions) interactions not observed with nucleic acids. Furthermore, PNA is achiral, providing it with the capability of adopting structural conformations the equivalent of which do not exist in the RNA/DNA realm.

The physico/chemical differences between PNA and DNA or RNA are also substantial. PNA binds to its complementary nucleic acid more rapidly than nucleic acid probes bind to the same target sequence. This behaviour is believed to be, at least partially, due to the fact that PNA lacks charge on its backbone. Additionally, recent publications demonstrate that the incorporation of positively charged groups into PNAs will improve the kinetics of hybridization (See: Iyer et al. J. Biol. Chem. (1995) 270, 14712-14717). Because it lacks charge on the backbone, the stability of the PNA/nucleic acid complex is higher than that of an analogous DNA/DNA or RNA/DNA complex. In certain situations, PNA will form highly stable triple helical complexes or form small loops through a process called "strand displacement". No equivalent strand displacement processes or structures are known in the DNA/RNA world.

In summary, because PNAs hybridize to nucleic acids with sequence specificity, PNAs are useful candidates for developing probe-based assays. However, PNA probes are not the equivalent of nucleic acid probes. Nonetheless, even under the most stringent conditions both the exact target sequence and a closely related sequence (e.g. a non-target sequence having a single point mutation (single base pair mismatch)) will often exhibit detectable interaction with a labelled nucleic acid or labelled PNA probe (See: Nielsen et al. Anti-Cancer Drug Design at p. 56-57 and Weiler et al. at p. 2798, second full paragraph). Any hybridization to a closely related non-target sequence will result in the generation of undesired background signal. Because the sequences are

so closely related, point mutations are some of the most difficult of all nucleic acid modifications to detect using a probe-based assay. Numerous diseases, such as sickle cell anemia and cystic fibrosis, are caused by a single point mutation of genomic nucleic acid. Consequently, any method, kits or compositions which could improve the specificity, sensitivity and reliability of probe-based assays would be useful in the detection, analysis and quantitation of DNA containing samples.

Sodium Bisulfite – A specific deamination method

Methods for treating nucleic acid with sodium bisufite can be found in a number of references including Frommer et al 1992, Proc Natl Acad Sci 89:1827-1831; Grigg and Clark 1994 BioAssays 16:431-436; Shapiro et al 1970, J Amer Chem Soc 92:422 to 423; Wataya and Hayatsu 1972, Biochemistry 11:3583 - 3588.

Detection Systems

Coating Magnetic beads

The PNA used for attachment to the magnetic beads can be modified in a number of ways. In this example, the PNA contained either a 5' or 3' amino group for the covalent attachment of the PNA to the beads using a hetero-bifunctional linker such as is used EDC. However, the PNA can also be modified with 5' groups such as biotin which can then be passively attached to magnetic beads modified with avidin or steptavidin groups.

Ten µL of carboxylate modified Magnabind™ beads (Pierce) or 100 µL of Dynabeads™ Streptavidin (Dynal) were transferred to a clean 1.5 mL tube and 90 µL of PBS solution added to the magnetic beads.

The beads were mixed then magnetised and the supernatant discarded. The beads were washed x2 in 100 µL of PBS per wash and finally resuspended in 90 µL of 50 mM MES buffer pH 4.5 or another buffer as determined by the manufactures' specifications.

One µL of 250 µM PNA (concentration dependant on the specific activity of the selected PNA as determined by oligonucleotide hybridisation experiments) is added to the sample and the tube vortexed and left at room temperature for 10-20 minutes.

Ten µL of a freshly prepared 25 mg/mL EDC solution (Pierce/Sigma) is then added, the sample vortexed and incubated at either room temperature or 4°C for up to 60 minutes.

The samples were then magnetised, the supernatant discarded and the beads, if necessary, be blocked by the addition of 100 µL either 0.25 M NaOH or 0.5 M Tris pH 8.0 for 10 minutes.

The beads were then washed x2 with PBS solution and finally resuspended in 100µl PBS solution.

Hybridisation using the magnetic beads

Ten µL of PNA coated Magnabind™ beads were transferred to a clean tube and 40 µL of either ExpressHyb™ buffer (Clontech) either neat or diluted 1:1 in distilled water or any other commercial or in-house hybridization buffer. The buffers may also contain either cationic/anionic or zwittergents at known concentration or other additives such as Heparin and poly amino acids.

Heat denatured sample of DNA 1-5 µL was then added to the above solution and the tubes vortexed and then incubated at 55°C or another temperature depending on the melting temperature of the chosen PNA for 20-60 minutes.

The samples were magnetised and the supernatant discarded and the beads washed x2 with 0.1XSSC/0.1%SDS at the hybridisation temperature from earlier step for 5 minutes per wash, magnetising the samples between washes.

Dual PNA capture

PNA#1 was coupled to a carboxylate modified magnetic bead via a N- or C-terminal amine of the PNA and washed to remove unbound PNA.

The PNA/bead complex is then hybridised to the target DNA in solution using appropriate hybridisation and washing conditions.

The target DNA was then released from the magnetic bead using appropriate methods and transferred to a tube containing a second PNA/magnetic beads complex targeted to the opposite end of the DNA molecule.

The second PNA/bead complex or oligo/bead complex was then hybridised to the target DNA in solution using appropriate hybridisation and washing conditions.

A third PNA or oligonucleotide complementary to the central region of the target DNA could be used as a detector molecule. This detector molecule can be labelled in a number of ways.

(i) The PNA or oligonucleotide can be directly labelled with a radioactive isotope such as P^{32} or I^{125} and then hybridised with the target DNA.

(ii) The PNA or oligonucleotide can be labelled with a fluorescent molecule such as Cy-3 or Cy-5 and then hybridised with the target DNA.

(iii) An amine modified PNA or oligonucleotide can be labelled in either of the above ways then coupled to a carboxylate modified microsphere of known size then the sphere washed to remove unbound labelled PNA or oligo. This bead complex can then be used to produce a signal amplification system for the detection of the specific DNA molecule.

(iv) The PNA or oligonucleotide can be attached to a dendrimer molecule either labelled with fluorescent or radioactive groups and this complex used to produce a signal amplification.

(v) The PNA or oligonucleotide labelled in any of the above ways and hybridised to the target DNA on a solid support can be released into solution using a single stranded specific nuclease such a mung bean nuclease or S1 nuclease. The released detector molecule can be read in a flow cytometer like device.

Preparation of radio-labelled detector spheres

A PNA or oligonucleotide molecule can be either 3' or 5' labelled with a molecule such as an amine group, thiol group or biotin.

The labelled molecule can also have a second label such as P^{32} or I^{125} incorporated at the opposite end of the molecule to the first label.

This dual labelled detector molecule can be covalently coupled to a carboxylate or modified latex bead for example of known size using a hetero-bifunctional linker such as EDC. Other suitable substrates can also be used depending on the assay.

The unbound molecules can then be removed by washing leaving a bead coated with large numbers of specific detector/signal amplification molecules.

These beads can then be hybridised with the DNA sample of interest to produce signal amplification.

Preparation of fluorescent labelled detector spheres

A PNA or oligonucleotide molecule can be either 3' or 5' labelled with a molecule such as an amine group, thiol group or biotin.

The labelled molecule can also have a second label such as Cy-3 or Cy-5 incorporated at the opposite end of the molecule to the first label.

This dual labelled detector molecule can now be covalently coupled to a carboxylate or modified latex bead of known size using a hetero-bifunctional linker such as EDC.

The unbound molecules can then be removed by washing leaving a bead coated with large numbers of specific detector/signal amplification molecules.

These beads can then be hybridised with the DNA sample of interest to produce signal amplification.

Preparation of enzyme labelled detector spheres

A PNA or oligonucleotide molecule can be either 3' or 5' labelled with a molecule such as an amine group or a thiol group.

The labelled molecule can also have a second label such as biotin or other molecules such as horse-radish peroxidase or alkaline phosphatase conjugated on via a hetero-bifunctional linker at the opposite end of the molecule to the first label.

This dual labelled detector molecule can now be covalently coupled to a carboxylate or modified latex bead of known size using a hetero-bifunctional linker such as EDC.

The unbound molecules can then be removed by washing leaving a bead coated with large numbers of specific detector/signal amplification molecules.

These beads can then be hybridised with the DNA sample of interest to produce signal amplification.

Signal amplification can then be achieved by binding of a molecule such as streptavidin or an enzymatic reaction involving a colorimetric substrate.

PNA oligomer combinations

In all of the above cases the initial hybridization event involved the use of magnetic beads coated with a PNA complimentary to the DNA of interest.

The second hybridisation event can involve any of the methods mentioned above.

This hybridisation reaction can be done with either a second PNA complimentary to the DNA of interest or an oligonucleotide or modified oligonucleotide complementary to the DNA of interest. As fluorescent beads of convenient size in these assays, carry $>10^6$ fluorochrome molecules and a single fluorescent bead can be detected readily, the method has the potential sensitivity to assay one or a few DNA molecules from one or a few cells.

Dendrimers and aptamers

Dendrimers are branched tree-like molecules that can be chemically synthesised in a controlled manner so that multiple layers can be generated that were labelled with specific molecules. They were synthesised stepwise from the centre to the periphery or visa-versa.

One of the most important parameters governing dendrimer structure and its generation is the number of branches generated at each step; this determines the number of repetitive steps required to build the desired molecule.

Dendrimers can be synthesised that contain radioactive labels such as I^{125} or P^{32} or fluorescent labels such as Cy-3 or Cy-5 to enhance signal amplification.

Alternatively dendrimers can be synthesised to contain carboxylate groups or any other reactive group that could be used to attach a modified PNA or DNA molecule.

RESULTS

Figure 1 and Figure 2 show examples of the method of the invention using sandwich PNA signal amplification using solid supports and magnetic beads, respectively. Although PNA is exemplified as the ligand in Figure 1 and Figure 2, it will be appreciated that other capture or detector ligands such as oligonucleotides can be used in these methods.

A solid support in the form of a microfilter well was provided and coated with N-oxysuccinimide to assist in the adhesion of PNA or other ligand to the well.

A first PNA which is complementary to a first part of the target nucleotide sequence is added to the well and attached to this solid support.

Bisulfite treated DNA is then added to the well and allowed to hybridise with the PNA to capture the target DNA which has hybridised to the PNA and subsequently bound to the well.

The well is then washed to remove the hybridisation solution and any non-hybridised DNA leaving only the hybridised DNA captured on the well.

Next a second PNA, which is complementary to a second part of the target nucleotide sequence is linked to microsphere beads having fluorescent labelling. The second linked PNA is then hybridised with the target DNA already bound to the well. The well is then washed to remove the unhybridized second PNA/microsphere complex leaving only the PNA/microsphere complex and fluorescent label associated with the target DNA sequence.

The fluorescence was then measured to determine the level of target DNA.

EXAMPLES

I. Detection of methylated promoter sequences of the GSTP1 zone

The promoter region of the *GSTP1* gene has been shown to be hypermethylated in prostate cancer (Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA, Bova GS, Hsieh W-S, Isaacs WB and Nelson WG. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies prostatic carcinogenesis. PNAS 91:11733-11737 (1994)). Bisulfite sequencing has defined region and specific CpG sites that were methylated in prostate cancer cells, but not in normal prostate (Millar DS, Ow KK, Paul CL, Russell PJ, Molloy PL and Clark SJ. Detailed methylation analysis of the glutathione S-transferase pi (*GSTP1*) gene in prostate cancer. Oncogene 18:1313-1324 (1999); Millar DS, Paul C, Molloy PL and SJ Clark (2000). A distinct sequence $(ATAAA)_n$ separates methylated and unmethylated domains at the 5' end of the *GSTP1* CpG island. J. Biol Chem. 275: 24893-24899). These studies define regions that provide good discrimination between DNA from normal tissues and DNA from prostate cancer cells, based on their different methylation pattern.

The sequence of part of this region of the GSTP1 gene is shown in Figure 3. Individual CpG sites were numbered above the sequence relative to the position of the transcription start site. Base numbers relative to the transcription start site were shown at the ends of lines of sequence. The sequence corresponds to bases 981 to 1160 of Genbank Accession No. M24485 (with inclusion of an additional CG). The number of each CpG site is indicated above

each site. Also indicated above CpG site -33 is the sequence of the polymorphic variant (p) which occurs in this region. The top line of each triplet shows the normal DNA sequence; the second line, B-U, shows the sequence that would arise following bisulfite treatment of DNA that contained no methylated cytosines (cytosines converted to uracils); the third line, B-C, shows the bisulfite-modified sequence produced if all cytosines at CpG sites were methylated. The position of PNAs #1 to #10 is shown under the sequence.

PNAs were synthesised that would hybridise to specific sites in the bisulfite-treated DNA as shown. Regions of sequence were chosen that contained a number of cytosines both within CpG sites (and potentially methylated) and not in CpG sites. PNAs were designed so that they will match perfectly if all cytosines at CpG sites in the DNA were methylated and hence had remained as cytosines and if all other cytosines had been efficiently converted to uracils. Thus, only properly bisulfite-converted, methylated DNA sequences should hybridise with the PNA probes under discriminating hybridisation conditions. The sequences of the ten PNA probes are shown below:

PNA#1 P-Linker- GAA ACA TCG CGA A-NH_2 SEQ ID NO:2

PNA#2 P-Linker- GAA ACA TCG CGA AAA-NH_2 SEQ ID NO:3

PNA#3 P-Linker- ATC GCC GCG CAA CTA A-NH_2 SEQ ID NO:4

PNA#4 P-Linker- AAA ACA TCA CAA AAA -NH_2 SEQ ID NO:5

PNA#5 P-Linker-ATC ACC ACA CAA CTA A-NH_2 SEQ ID NO:6

PNA#6 P-Linker-CTA ACG CGC CGA AAC-NH_2 SEQ ID NO:7

PNA#7 P-Linker-CCA CTA CAA TCC CA-NH_2 SEQ ID NO:8

PNA#8 P-Linker-CAC CAC ACA ACT-NH_2 SEQ ID NO:9

PNA#9 P-Linker-GCA ACT AAG CAA CG-NH_2 SEQ ID NO:10

PNA#10 P-Linker-GCA ACG AAC TAA CG-NH_2 SEQ ID NO:11

Example (i)

Using the approach shown in Figure 1 PNA#2 was coupled to wells of a microtitre tray. One to 2 μg of DNA from the prostate cancer cell lines LNCaP, PC-3-M and DU145 was treated with bisulfite as described (Clark SJ, Harrison J, Paul CL and Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 22: 2990-2997 (1994)) and resuspended in 100 μL. DNAs were diluted 1:100 with ExpressHyb buffer and 100 μL samples added to wells for hybridisation. One μg of salmon sperm DNA was used in control wells. After hybridisation for 2 hr and washing steps hybridisation was carried out with PNA#3 coupled to either 0.5 μM or 0.1 μM fluorospheres.

Fluorescent signals were for both LNCaP and PC-3-M DNAs in comparison to DU145 and the negative control salmon sperm DNAs (Figure 4) after background subtraction. In all cases, a higher signal was seen when the PNA#3 was coupled to the larger diameter (0.5 μM) spheres.

Genomic sequencing has shown that the GSTP1 gene is heavily methylated in LNCaP DNA and significantly methylated in PC-3-M DNA. DNA from the DU145 cell line was shown however to be under methylated (<10%) across the region targeted by the PNA probes used (Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA, Bova GS, Hsieh W-S, Isaacs WB and Nelson WG. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies prostatic carcinogenesis. PNAS 91:11733-11737 (1994)). The assay is thus able to distinguish between methylated and unmethylated DNAs following bisulfite treatment.

Example (ii)

Following the protocol used in Example (i) bisulfite-treated DNA from the LNCaP and PC-3-M cell lines as well as DNA extracted from normal prostate tissue and two samples of prostate cancer tissue were assayed. The normal DNA had been shown to be unmethylated and the cancer DNA samples methylated in the target region. The data (Figure 5) demonstrate that DNA from normal and cancerous prostate tissue can be distinguished.

Example (iii)

In Figure 6 the effects of the level of PNA coated on the wells and of the concentration of the target DNA population were shown. Wells were coated with 0.1, 1 or 10 nmoles of PNA (10 nmoles in previous experiments) and serial dilutions of bisulfite-treated LNCaP DNA. The amounts correspond to 10 ng, 1 ng, 200 pg and 100 pg of DNA prior to bisulfite treatment.

Sensitivity of detection was greatest with 10 pmoles of PNA attached to the wells, with LNCaP DNA corresponding to an input of 100 pg being detectably above the salmon sperm control. Background signals from control salmon sperm DNA also increased as a function of the amount of PNA on the well. PCR amplifications using methylation specific PCR primers were also carried out on the same bisulfite-treated LNCaP DNA samples. The primers used selectively amplify bisulfite-treated DNA corresponding to the GSTP1 promoter methylated at target CpG sites. Primers used were CGPS-5 and CGPS-8 for first round amplification and CGPS-11 and CGPS-12 for second round amplification under conditions as described in PCT/AU99/00306. The data showed that detection by the PNA/fluorescent bead capture assay was at least as sensitive as detection by methylation specific PCR.

II. Detection of Captured DNA using a Fluorescent Dye

Methylated GSTP1 promoter sequences were detected using bisulfite-treated LNCaP DNA and the single-stranded DNA-binding dye Oligreen (Molecular Probes catalogue number 07582). The Oligreen will bind to any hybridised (captured) DNA remaining after washing steps but will not bind to the PNA probes attached to the wells.

PNA #2 and #3 were coupled to wells of a microtitre plate (1 pMole per well) and 1 μg of bisulfite-treated LNCaP DNA hybridised as above; 1 μg of salmon sperm DNA was used as the control. Hybridisation was done using either ExpressHyb Buffer (Clontech) or GDA hybridisation buffer (0.75 M NaCl, 0.17 M sodium phosphate, 0.1% (w/v) sodium pyrophosphate, 0.15 M Tris, pH 7.5, 2%

sodium dodecyl sulphate, 100 μg/mL salmon sperm DNA, 5X Denhardt's solution [0.1% ficoll, 0.1% bovine serum albumin, 0.1% polyvinylpyrrolidone]). After three washes in 150 μL of water, captured DNA was incubated with a solution of the dye Oligreen that only fluoresces when bound to single stranded DNA. Oligreen stock solution as supplied by the manufacturer was diluted 1:20 in phosphate buffered saline containing 1 mM EDTA and 100 μL added per well. After 5 min incubation fluorescence was read using a 500 nm excitation filter and a 520 nm emission filter.

For both PNA#2 and PNA#3 optimal results were obtained using the GDA hybridisation buffer (Figure 7). In both cases, fluorescence from methylated LNCaP DNA could be detected above the level of control salmon sperm DNA.

III. Detection of methylated DNA using microspheres

Methodology

Referring to Figure 1 and Figure 2 approaches for detection of methylated DNA using microspheres is shown.

Coating Microtitre wells with capture PNA

(i) The capture PNA#2 (0.0-100 pM per well) in 50 mM Phosphate buffer, 1 mM EDTA pH 8.5 (100 μL) was used to coat N-oxysuccinimide-coated microtitre wells (Costar Cat#2498) for 16-24 hours @ 4°C.

(ii) Plates were washed with 100 μL of 50 mM Phosphate buffer, 1 mM EDTA pH 8.5.

(iii) 150 μL of 3% BSA, 50 mM Phosphate buffer, 1 mM EDTA pH 8.5 was added to each well and the plates left @ 4°C until required.

Coating the Fluorospheres with detection PNA

(i) Fluorospheres (Molecular Probes) were sonicated five times for 5 seconds to break up any aggregated material.

(ii) The detection probe PNA#3 was diluted in a range from 300 pM to 0.3 pM in 250 μL of sonicated 50 mM 2[N-morpholino] ethanesulphonic acid (MES) pH 6.0 and 250 μL of sonicated fluorospheres added and the solution left at room temperature for 30 minutes.

(iii) 0.5 mg of 1-ethyl-3[3 dimethylamine propyl] carbodiimide [EDAC], Sigma Cat #E1769, was added to the sample and the sample left 4-6 hours at room temperature in the dark then incubated 16 hours at 4°C.

(iv) 55 μL of 1M glycine was added to the beads and the beads left at room temperature for 2 hours.

(v) The beads were centrifuged for 5-20 minutes (dependant on size of beads, generally 0.5 μM beads required 5 mins while 0.1 μM beads required 20 minutes) at 14,000 rpm in a bench top centrifuge and the supernatant discarded.

(vi) Beads were washed twice with 500 μL of PBS/1% BSA with centrifugation as before between wash steps.

(vii) The beads were then resuspended in 200 μL of PBS/1% BSA and stored at 4°C in the dark until required.

(viii) Variation of the number of PNAs bound to the beads can be used to optimise sensitivity and minimise background levels.

Hybridisation of DNA

(i) Either control salmon sperm DNA or DNA that was bisulfite treated as in Clark *et al* (Clark SJ, Harrison J, Paul CL and Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 22: 2990-2997 (1994)) was hybridised with PNAs coupled to microtitre wells then added to per well.

(i) DNA samples were mixed with 100 μL of ExpressHyb™ buffer (Clontech), added to the wells and the plate covered with cling film

or the wells overlayed with mineral oil (Sigma) for longer incubations and the samples incubated at between 45-60°C for between 1-16 hours.

(iii) Wells were then washed twice with 150 µL of 2X SSC/0.1%SDS @ 45-60°C for 5-10 minutes per wash.

(iv) The wells were further washed with 150 µL of 0.1X SSC/0.1%SDS @ 45-60°C for 5-10 minutes and the wash solution discarded.

(i) The PNA/fluorospheres were diluted 1/100 in ExpressHyb™ buffer (Clontech) and 100 µL of samples added to the wells. The plates were covered with cling film or the wells overlayed with mineral oil (Sigma) for longer incubations and the samples incubated @ between 45-60°C for between 1-16 hours.

(vi) Wells were then washed twice with 150 µL of 2X SSC/0.1%SDS at 45-60°C for 5-10 minutes per wash.

(vii) The wells were further washed with 150 µL of 0.1X SSC/0.1%SDS at 45-60°C for 5-10 minutes and the wash solution discarded.

(viii) Finally the fluorescent intensity of each well was measured at the appropriate excitation/emission wave-length for the particular bead (500/520 for yellow beads) in a Victor II fluorescent plate reader.

(ix) Background values measured in wells to which no PNA had been attached were subtracted from all readings.

IV. Method for the production of in-house coated radiolabelled beads

(i) A specific oligonucleotide (or PNA) is synthesised against the target DNA region of interest. This oligonucleotide contains a 3' amine group synthesised using standard chemistry (Sigma Genosys).

(ii) The oligonucleotide (or PNA) is then 5' kinased using gamma P^{32}dATP as follows:

- Oligonucleotide (20 ng/µL) 1 µL
- X10 PNK buffer 1 µL

- T4 PNK 1 µL
- Gamma P^{32}dATP 2 µL
- Sterile water 5 µL

(iii) The sample is then incubated at 37°C for 1 hour then heated to 95°C for 5 minutes to inactivate the enzyme.

(iv) 0.1 µM carboxylate modified fluorescent beads (Molecular Probes Cat# F-8803) are diluted 1/10,000, 1/100,000 and 1/1,000,000 in sterile water then the kinased oligonucleotide coupled to the beads as follows:

- Beads 1 µL
- Labelled oligo 3 µL
- 50 mM MES pH 8.0 5 µL
- 10 mg/mL EDC (Pierce) 2 µL

(v) The beads are then incubated @ room temperature for 1 hour to allow the kinased oligonucleotide to attach to the beads via the 3' amine.

(vi) The beads are then spun in a microfuge at full speed for 15 minutes to sediment the coated beads.

(vii) The supernatant is removed and the beads washed with 100 µL of PBS solution and spun as above.

(viii) The supernatant is removed and the beads resuspended in 50 µL of PBS.

(ix) The CPM of the coated beads is then measured in a standard scintillation counter using the Cerenkov counting protocol. The beads with the highest activity are then used as a detector system in the assay

The idea behind this protocol is to produce the smallest number of beads with the highest specific activity, so that only a few beads are needed to bind to the target sequence in order to generate a detectable signal.

V. Urea and Methyoxyamine conditions of use

(i) Typically 2 µg of genomic DNA is restriction digested with an appropriate enzyme (as determined by the target DNA sequence) under the manufacturers conditions for at least 2 hours in a final volume of 20 µL.

(ii) 2.2 µL of freshly prepared 3 M NaOH (6g in 50 mL H_2O) is added to the DNA and the sample incubated @ 37°C for 15 minutes.

(iii) 6.24 g of urea is added to 10 mL of sterile distilled water and the solution mixed until gently until the urea has dissolved.

(iv) 7.6 g of sodium metabisulphite (BDH Analar™) is then added and again the solution mixed gently until the bisulphite had dissolved.

(v) The pH of the reagent is then adjusted to 5.0 with 10 M NaOH and the volume of the reagent made to 20 mL with sterile water.

(vi) 208 µL of the reagent is then added to the digested denatured genomic DNA sample.

(vii) 12 µL of 10-100 mM quinol is added the solution mixed and overlayed with mineral oil and incubated for 16 hours at 55°C in the dark.

(viii) The mineral oil is then removed and the DNA purified using the Promega Wizard ™ DNA purification system according to the manufacturers instructions.

(ix) The DNA is eluted from the column with 50 µL of sterile water then 5.5 µL of 3 M NaOH added and the sample incubated at 37°C for 15 minutes.

(x) At this stage 1/10 volume of methoxyamine (Sigma) from 1-100 mM can be added and incubated with the NaOH as an agent to minimise the nicking of the bisulphite treated DNA.

(xi) In addition tRNA can or glycogen may be added at this stage to help precipitate the DNA

(xii) The DNA is then precipitated by the addition of 33.5 µL of 5 M $NH_4OAcetate$ pH 7.0 and 330 µL of 100% ethanol.

(xiii) The samples are incubated at -20°C for at least 1 hour then spun down in a microcentrifuge at full speed for 15 minutes

(xiv) The pellet is then air dried for 5-10 minutes and the DNA resuspended in 10 mM Tris/0.1 mM EDTA pH 8.0 in volumes ranging from 10-100 µL dependant on the downstream processing of the modified DNA.

Variations on the above protocol are set out below.

If very small quantities of DNA or micro-dissected cells are to be bisulfite treated this can be done in a number of ways.

- Restriction digestion can be omitted.
- Urea can be omitted
- Glycogen or tRNA or a combination of both can be added at steps (iv), (viii) and (x).
- The bisulfite reaction can be done by encapsulating the DNA to be modified in agarose bead, and the entire reaction carried out while the DNA is in the bead.
- The time of the reaction with the bisulphite can be reduced from 16 hours to as little as 1 hour but more usually 4 hours.

SUMMARY

The methods of the present invention can be applied for the detection of any DNA using one ligand (preferably an oligonucleotide or PNA) bound to a solid support and one coupled to a microsphere. Natural oligonucleotides or PNAs may be used, but PNAs were preferred because of their specificity and rate of hybridisation.

In one particular adaptation, the methods of the invention can be used to distinguish the presence of methylated cytosines in DNA that has been treated with sodium bisulfite. The specificity of hybridisation can be used to discriminate

against molecules that have not reacted completely with bisulfite (one or more cytosines not converted to uracil) as well as distinguishing between methylated cytosines at CpG sites (which remain as cytosines) and unmethylated CpG sites where the cytosine is converted to uracil.

In another adaptation the methods of the invention can be used to discriminate against DNA whose cytosines have not reacted completely with bisulfite reagent to convert them to uracils because they happen to carry a methyl group in the 5' position.

As treatment with bisulfite changes the sequence of the DNA by converting all cytosines (but not 5-methyl cytosines) to uracils, specific PNA's can be made which recognise a region having 5 methyl cytosines but which will not recognise the same sequence which happens to have no 5-methyl cytosines.

The methods of the invention can also be applied to the discrimination of mutant alleles of a gene where the sequence of one or both of the oligonucleotides or PNAs will match perfectly with one allele but mismatch with the other.

Figure 7 demonstrates the high sensitivity of the method of the invention showing sensitivity similar to that achieved using PCR techniques.

The method of the invention has numerous applications as previously described including particular use in devising multiple array chips for rapid detection of the methylation status of bulk DNA samples.

Figure 8 and Figure 9 show radioactive data of methylated molecules and unmethylated molecules indicating the sensitivity and specificity of the present invention. As can be seen from the results, the method is capable of distinguishing 1% methylation or unmethylation in a background of 99% unmethylated and 99% methylated molecules, respectively.

Although prostate cancer-related gene was used as an example of the use of the present invention, it will be appreciated that the methods are applicable for many other states and conditions where different methylation states have been found to play a role in disease or altered state of cells. Examples of just some genes affected by CpG island promoter methylation are shown in Table 3. The

present invention is clearly applicable for the detection or measurement of such methylation states and many others.

Table 3. Examples of genes affected by CpG or CNG island promoter methylation

Gene	Location	Cancer	Aging	Comments
APC	5q21	Colon, gastric, oesophageal	No	
BRACI	17q21	Breast, ovarian	No	
Calcitonin	11p15	Colon, lung, haematological	No	One of the first to be found methylated in cancer
E-cadherin	16q22.1	Breast, gastric, thyroid, SCC, leukemia, liver	No	
Estrogen Receptor	6q25.1	Colon, liver, heart, breast, lung	Yes	Good correlation between methylation and loss of expression
H19	11p15.5	Wilms tumour	No	Imprinted gene
HIC1	19p13.3	Prostate, breast, brain, lung	Yes	Candidate tumour suppressor
IGF2	11p15.5	Colon, AML	Yes	Has large CpG island
MDGI	1p33-35	Breast	No	
MGMT	10q26	Brain, colon, lung, breast	No	
MYOD1	11p15.4	Colon, breast, bladder, lung	Yes	

Table 3. cont'd Examples of genes affected by CpG or CNG island promoter methylation

Gene	Location	Cancer	Aging	Comments
N33	8p22	Colon, prostate, brain	Yes	Oligo-saccharyl-transferase
P15	9q21	Leukemia, lung, colon	No	
P16	9q21	Lung, colon, lymphoma, bladder, and more	No	Methylation occurs as frequent as deletions or other mutations
TIMP3	22q12.1	Brain, kidney	No	
WT1	11p13	Breast, colon, Wilms tumour	No	

Figure 10 shows an example of the methylation pattern of GSTP1 in prostate cancer. As can be seen, only subtle changes in the methylation state of this gene region have been implicated in this cancer state. The ability to detect such changes by the methods according to the present invention is a powerful tool for the early detection of cancer and other altered states in cells as well as determining the affect of therapeutic and other agents on cells and tissue.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments were, therefore, to be considered in all respects as illustrative and not restrictive.

Claims:

1. A method for detecting presence of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing to the treated sample a detector ligand in the form of a peptide nucleic acid (PNA) molecule capable of binding to a target region of DNA and allowing sufficient time for a detector ligand to bind to a target DNA; and

(c) measuring binding of the detector ligand to DNA in the sample to determine the presence of the target DNA in a sample, wherein the method is carried out without a DNA amplification step.

2. The method according to claim 1 wherein the sample is selected from the group consisting of biological, tissue, environmental, and microbiological.

3. The method according to claim 2 wherein the biological sample is selected from the group consisting of blood, urine, faeces, semen, cerebrospinal fluid, and cells.

4. The method according to claim 2 wherein the tissue sample is selected from the group consisting of brain, colon, urogenital, lung, renal, hematopoietic, breast, thymus, testis, ovary, and uterus.

5. The method according to claim 4 wherein the tissue sample is selected from the group consisting of brain, colorectal, and prostate.

6. The method according to claim 5 wherein the tissue sample is prostate.

AMENDED SHEET
IPEA/AU

7. The method according to claim 2 wherein the microbiological sample is selected from the group consisting of virus, viroid, bacteria, yeast, fungi, and protozoa.

8. The method according to any one of claims 1 to 7 wherein the modifying agent is capable of modifying unmethylated cytosine but not methylated cytosine.

9. The method according to claim 8 wherein the modifying agent is selected from the group consisting of bisulfite, acetate and citrate.

10. The method according to claim 9 wherein the modifying agent is sodium bisulfite and cytosine is modified to uracil.

11. The method according to claim 10 further including additives selected from the group consisting of urea, methoxyamine and mixtures thereof.

12. The method according to any one of claims 1 to 11 wherein the PNA molecule is from 5 to 40 bases in length.

13. The method according to any one of claim 1 to 12 wherein the PNA molecule is directed to a CpG- or CNG-containing region of DNA.

14. The method according to claim 13 wherein the CpG- or CNG-containing region of DNA is in a regulatory region of a gene or an enhancer of any regulatory element or region selected from the group consisting of promoter, enhancer, oncogene, or other regulatory element which activity is altered by environmental factors including chemicals, toxins, drugs, radiation, synthetic or natural compounds, and microorganisms or other infectious agents including viruses, bacteria, yeast, fungi, protozoa, and prions.

AMENDED SHEET
IPEA/AU

15. The method according to any one of claims 1 to 14 wherein the detector ligand contains a detectable label and the binding of the ligand to target DNA is detected by measuring the presence and/or amount of the detectable label associated with the DNA.

16. The method according to claim 15 wherein the detectable label is selected from the group consisting of fluorescence, radioactivity, enzyme, hapten and dendrimer.

17. A method for estimating extent of methylation of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing to the treated sample a detector ligand in the form of a peptide nucleic acid (PNA) molecule capable of distinguishing between methylated and unmethylated cytosine of DNA and allowing sufficient time for a detector ligand to bind to a target DNA; and

(c) detecting binding of the detector ligand to DNA in the sample such that the degree or amount of binding is indicative of the extent of methylation of the target DNA, wherein the method is carried out without a DNA amplification step.

18. The method according to claim 17 wherein the sample is selected from the group consisting of biological, tissue, environmental, and microbiological.

19. The method according to claim 18 wherein the biological sample is selected from the group consisting of blood, urine, faeces, semen, cerebrospinal fluid, and cells.

AMENDED SHEET
IPEA/AU

20. The method according to claim 19 wherein the tissue sample is selected from the group consisting of brain, colon, urogenital, lung, renal, hematopoietic, breast, thymus, testis, ovary, and uterus.

21. The method according to claim 20 wherein the tissue sample is selected from the group consisting of brain, colorectal, and prostate.

22. The method according to claim 21 wherein the tissue sample is prostate.

23. The method according to claim 18 wherein the microbiological sample is selected from the group consisting of virus, viroid, bacteria, yeast, fungi, and protozoa.

24. The method according to any one of claims 17 to 23 wherein the modifying agent is capable of modifying unmethylated cytosine but not methylated cytosine.

25. The method according to claim 24 wherein the modifying agent is selected from the group consisting of bisulfite, acetate and citrate.

26. The method according to claim 25 wherein the modifying agent is sodium bisulfite and cytosine is modified to uracil.

27. The method according to claim 26 further including additives selected from the group consisting of urea, methoxyamine and mixtures thereof.

28. The method according to any one of claims 17 to 27 wherein the PNA molecule is from 5 to 40 bases in length.

AMENDED SHEET
IPEA/AU

29. The method according to any one of claims claim 17 to 28 wherein the PNA molecule is directed to a CpG- or CNG-containing region of DNA.

30. The method according to claim 29 wherein the CpG- or CNG-containing region of DNA is in a regulatory region of a gene or an enhancer of any regulatory element or region selected from the group consisting of promoter, enhancer, oncogene, or other regulatory element which activity is altered by environmental factors including chemicals, toxins, drugs, radiation, synthetic or natural compounds, and microorganisms or other infectious agents including viruses, bacteria, yeast, fungi, protozoa, and prions.

31. The method according to any one of claims 17 to 30 wherein the detector ligand contains a detectable label and the binding of the ligand to target DNA is detected by measuring the presence and/or amount of the detectable label associated with the DNA.

32. The method according to claim 31 wherein the detectable label is selected from the group consisting of fluorescence, radioactivity, enzyme, hapten and dendrimer.

33. The method according to any one of claims 17 to 32 wherein two different detector ligands are used, wherein a first ligand being capable of binding to a region of DNA that contains one or more methylated cytosines and a second ligand being capable of binding to a corresponding region of DNA that contains no methylated cytosines, wherein at least one of the detector ligands is a PNA molecule.

34. The method according to claim 33 wherein the two ligands are added to the same treated sample and the binding of each ligand is detected in the one treated sample.

AMENDED SHEET
IPEA/AU

35. The method according to claim 33 wherein each ligand is added to a separate assay and the binding of each ligand is detected in each assay and the binding of the two ligands is compared.

36. A method for detecting the presence of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing a support to which is bound a capture ligand capable of recognising a first part of a target DNA sequence;

(c) contacting the support with the treated sample for sufficient time to allow DNA to bind to a capture ligand such that target DNA in the sample binds to the support via the capture ligand;

(d) contacting the support with a detector ligand in the form of a peptide nucleic acid (PNA) molecule capable of recognising a second part of the target DNA sequence and allowing sufficient time for a detector ligand to bind to a target DNA; and

(e) measuring binding of the detector ligand to DNA bound to the support to determine the presence of the target DNA in the sample, wherein the method is carried out without a DNA amplification step.

37. The method according to claim 36 wherein the sample is selected from the group consisting of biological, tissue, environmental, and microbiological.

38. The method according to claim 37 wherein the biological sample is selected from the group consisting of blood, urine, faeces, semen, cerebrospinal fluid, and cells.

39. The method according to claim 37 wherein the tissue sample is selected from the group consisting of brain, colon, urogenital, lung, renal, hematopoietic, breast, thymus, testis, ovary, and uterus.

AMENDED SHEET
IPEA/AU

40. The method according to claim 39 wherein the tissue sample is selected from the group consisting of brain, colorectal and prostate.

41. The method according to claim 40 wherein the tissue sample is prostate.

42. The method according to claim 37 wherein the microbiological sample is selected from the group consisting of virus, viroid, bacteria, yeast, fungi, and protozoa.

43. The method according to any one of claims 36 to 42 wherein the modifying agent is capable of modifying unmethylated cytosine but not methylated cytosine.

44. The method according to claim 43 wherein the modifying agent is selected from the group consisting of bisulfite, acetate and citrate.

45. The method according to claim 44 wherein the modifying agent is sodium bisulfite and cytosine is modified to uracil.

46. The method according to claim 45 further including additives selected from the group consisting of urea, methoxyamine, and mixtures thereof.

47. The method according to any one of claims 36 to 46 wherein the support is selected from the group consisting of glass, polymer including cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene, plastic materials, fluorescent beads, magnetic beads, synthetic or natural membranes, latex beads, column supports, beads or slides, nanotubes, fibres, and organic or inorganic solid supports.

AMENDED SHEET
IPEA/AU

48. The method according to claim 47 wherein the support is a magnetic bead or a fluorescent bead.

49. The method according to any one claims 36 to 47 wherein step (b) comprises a plurality capture ligands arrayed on a solid support.

50. The method according to claim 49 wherein the array contains multiple copies of the same ligand so as to capture the same target DNA on the array.

51. The method according to claim 49 wherein the array contains a plurality of different ligands targeted to different DNA so as to capture a plurality of target DNA molecules on the array.

52. The method according to any one of claims 49 to 51 wherein the array contains from 10 to 10000 capture ligands.

53. The method according to claim 52 wherein the array has less than 500 capture ligands.

54. The method according to any one of claims 36 to 53 wherein the capture ligand is selected from the group consisting of PNA molecule, oligonucleotide, modified oligonucleotide, single stranded DNA, RNA, aptamer, antibody, protein, peptide, a combination thereof, and chimeric versions thereof.

55. The method according to claim 54 wherein the capture ligand is a PNA molecule or an oligonucleotide molecule.

56. The method according to claim 55 wherein the capture ligand is a PNA molecule.

AMENDED SHEET
IPEA/AU

57. The method according to claim 56 wherein the capture PNA molecule is from 5 to 40 bases in length.

58. The method according to any one of claims 36 to 57 wherein the detector PNA molecule is from 5 to 40 bases in length.

59. The method according to any one of claims 36 to 58 wherein the PNA molecule is directed to a CpG- or CNG-containing region of DNA.

60. The method according to claim 59 wherein the CpG- or CNG-containing region of DNA is in a regulatory region of a gene or an enhancer of any regulatory element or region selected from the group consisting of promoter, enhancer, oncogene, or other regulatory element which activity is altered by environmental factors including chemicals, toxins, drugs, radiation, synthetic or natural compounds, and microorganisms or other infectious agents including viruses, bacteria, yeast, fungi, protozoa, and prions.

61. The method according to any one of claims 36 to 60 wherein the detector ligand contains a detectable label and the binding of the ligand to target DNA is detected by measuring the presence and/or amount of the detectable label.

62. The method according to claim 61 wherein the detectable label is selected from the group consisting of fluorescence, radioactivity, enzyme, hapten and dendrimer.

63. The method according to any one of claims 36 to 62 wherein two different detector ligands are used, wherein a first ligand being capable of binding to a region of DNA that contains one or more methylated cytosines and a second ligand being capable of binding to a corresponding region of DNA that contains no methylated cytosines, wherein at least one of the detector ligands

AMENDED SHEET
IPEA/AU

is a PNA molecule.

64. The method according to claim 63 wherein the two detector ligands are added to the same treated sample and the binding of each ligand is detected in the one treated sample.

65. The method according to claim 63 wherein each detector ligand is added to a separate assay and the binding of each ligand is detected in each assay and the binding of the two ligands is compared.

66. A method for estimating extent of methylation of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with an agent that modifies unmethylated cytosine;

(b) providing a support to which is bound a capture ligand which is capable of recognising a first part of a target DNA sequence;

(c) contacting the support with the treated sample for sufficient time to allow DNA to bind to a capture ligand such that target DNA in the sample binds to the support via the capture ligand;

(d) contacting the support with a detector ligand in the form of a peptide nucleic acid (PNA) molecule capable of distinguishing between methylated and unmethylated cytosine of DNA such that the detector ligand binds to any target DNA on the support; and

(e) detecting binding of the detector ligand to the support such that the degree or amount of binding is indicative of the extent of methylation of the target DNA, wherein the method is carried out without a DNA amplification step.

67. The method according to claim 66 wherein the sample is selected from the group consisting of biological, tissue, environmental, and microbiological.

AMENDED SHEET
IPEA/AU

68. The method according to claim 67 wherein the biological sample is selected from the group consisting of blood, urine, faeces, semen, cerebrospinal fluid, and cells.

69. The method according to claim 67 wherein the tissue sample is selected from the group consisting of brain, colon, urogenital, lung, renal, hematopoietic, breast, thymus, testis, ovary, and uterus.

70. The method according to claim 69 wherein the tissue sample is selected from the group consisting of brain, colorectal, and prostate.

71. The method according to claim 70 wherein the tissue sample is prostate.

72. The method according to claim 67 wherein the microbiological sample is selected from the group consisting of virus, viroid, bacteria, yeast, fungi, and protozoa.

73. The method according to any one of claims 66 to 72 wherein the modifying agent is capable of modifying unmethylated cytosine but not methylated cytosine.

74. The method according to claim 73 wherein the modifying agent is selected from the group consisting of bisulfite, acetate and citrate.

75. The method according to claim 74 wherein the modifying agent is sodium bisulfite and cytosine is modified to uracil.

76. The method according to claim 75 further including additives selected from the group consisting of urea, methoxyamine, and mixtures thereof.

AMENDED SHEET
IPEA/AU

77. The method according to any one of claims 66 to 76 wherein the support is selected from the group consisting of glass, polymer including cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene, plastic materials, fluorescent beads, magnetic beads, synthetic or natural membranes, latex beads, column supports, beads or slides, nanotubes, fibres, and organic or inorganic solid supports.

78. The method according to claim 77 wherein the support is a magnetic bead or a fluorescent bead.

79. The method according to any one claims 66 to 78 wherein step (b) comprises a plurality capture ligands arrayed on a solid support.

80. The method according to claim 79 wherein the array contains multiple copies of the same ligand so as to capture the same target DNA on the array.

81. The method according to claim 80 wherein the array contains a plurality of different ligands targeted to different DNA so as to capture a plurality of target DNA molecules on the array.

82. The method according to any one of claims 66 to 81 wherein the array contains from 10 to 10,000 capture ligands.

83. The method according to claim 82 wherein the array has less than 500 capture ligands.

84. The method according to any one of claims 66 to 83 wherein the capture ligand is selected from the group consisting of PNA molecule, oligonucleotide, modified oligonucleotide, single stranded DNA, RNA, aptamer, antibody, protein, peptide, a combination thereof, and chimeric versions thereof.

AMENDED SHEET
IPEA/AU

85. The method according to claim 84 wherein the capture ligand is a PNA molecule or an oligonucleotide molecule.

86. The method according to claim 85 wherein the capture ligand is a PNA molecule.

87. The method according to claim 86 wherein the PNA molecule is from 5 to 40 bases in length.

88. The method according to any one of claims 66 to 87 wherein the detector PNA molecule is from 5 to 40 bases in length.

89. The method according to any one of claims 66 to 88 wherein the PNA molecule is directed to a CpG- or CNG-containing region of DNA.

90. The method according to claim 89 wherein the CpG- or CNG-containing region of DNA is in a regulatory region of a gene or an enhancer of any regulatory element or region selected from the group consisting of promoter, enhancer, oncogene, or other regulatory element which activity is altered by environmental factors including chemicals, toxins, drugs, radiation, synthetic or natural compounds, and microorganisms or other infectious agents including viruses, bacteria, yeast, fungi, protozoa, and prions.

91. The method according to any one of claims 66 to 90 wherein the detector ligand contains a detectable label and the binding of the ligand to target DNA is detected by measuring the presence and/or amount of the detectable label.

92. The method according to claim 91 wherein the detectable label is selected from the group consisting of fluorescence, radioactivity, enzyme, hapten and dendrimer.

AMENDED SHEET
IPEA/AU

93. The method according to any one of claims 66 to 92 wherein two different detector ligands are used, wherein a first ligand being capable of binding to a region of DNA that contains one or more methylated cytosines and a second ligand being capable of binding to a corresponding region of DNA that contains no methylated cytosines, wherein at least one of the detector ligands is a PNA molecule.

94. The method according to claim 93 wherein the two detector ligands are added to the same treated sample and the binding of each ligand is detected in the one treated sample.

95. The method according to claim 93 wherein each detector ligand is added to a separate assay and the binding of each ligand is detected in each assay and the binding of the two ligands is compared.

96. A method for detecting a methylated CpG- or CNG-containing DNA, the method comprising:

(a) treating a sample containing DNA with bisulfite to modify unmethylated cytosine to uracil in the DNA;

(b) providing to the treated sample a detector in the form of a peptide nucleic acid (PNA) ligand capable of distinguishing between methylated and unmethylated cytosine of DNA; and

(c) detecting the methylated DNA based on the presence or absence of binding of the detector PNA ligand, wherein the method is carried out without a DNA amplification step.

97. The method according to claim 96 wherein the detector PNA ligand is capable of binding to a methylated CpG- or CNG-containing DNA but not to a corresponding unmethylated CpG- or CNG-containing DNA. and binding of the PNA ligand to DNA is indicative of methylation of the DNA.

AMENDED SHEET
IPEA/AU

98. A method for estimating extent of methylation of a target DNA in a sample, the method comprising:

(a) treating a sample containing DNA with bisulfite to modify unmethylated cytosine to uracil;

(b) providing a solid support in the form of a magnetic bead to which is bound a capture peptide nucleic acid (PNA) ligand or oligonucleotide ligand which is capable of recognising a first part of a target DNA sequence;

(c) contacting the support with the treated sample suspected of containing the target DNA such that target DNA in the sample binds to the support via the capture ligand;

(d) contacting the support with a detector PNA ligand capable of distinguishing between methylated and unmethylated cytosine of DNA; and

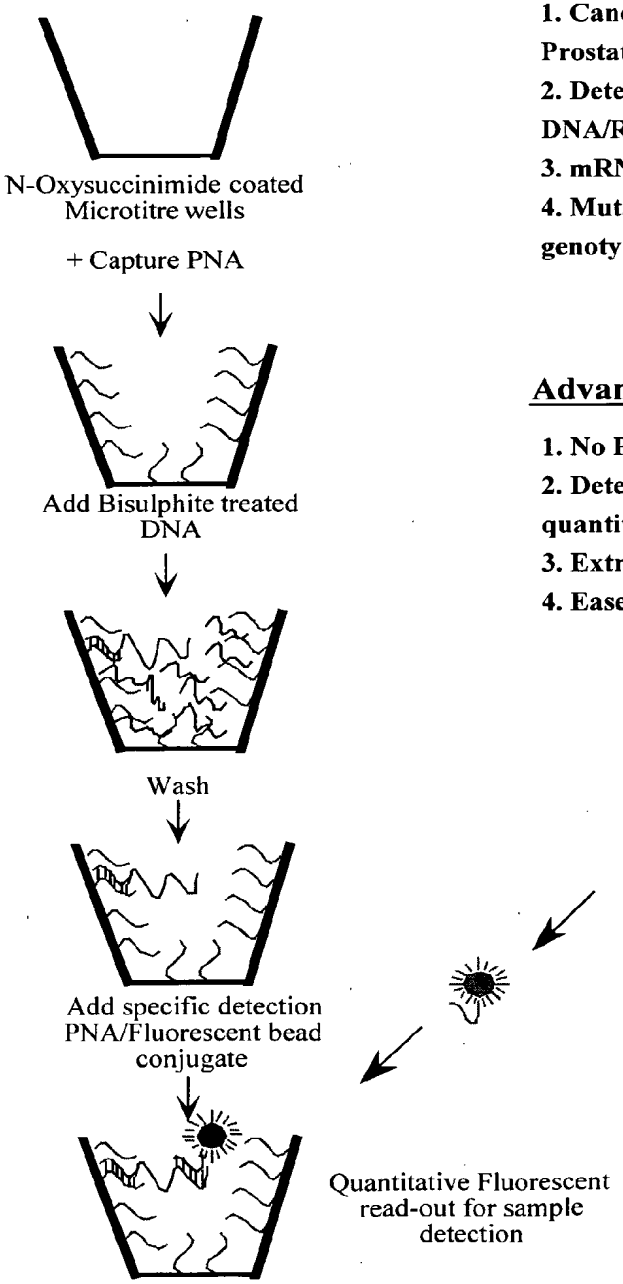
(e) determining the extent of methylation of the DNA bound to the support by measuring the amount of bound detector ligand, wherein the method is carried out without a DNA amplification step.

AMENDED SHEET
IPEA/AU

Figure 1

Sandwich PNA Signal Amplification Technology

Methodology



Applications

1. Cancer Diagnostics e.g Prostate.
2. Detection of single stranded DNA/RNA viruses e.g Influenza.
3. mRNA Micro-arrays.
4. Mutational Analysis e.g genotyping.

Advantages

1. No PCR Amplification step.
2. Detection would be quantitative.
3. Extremely specific.
4. Ease of automation.

Figure 2 Sandwich PNA signal amplification technology using magnetic beads

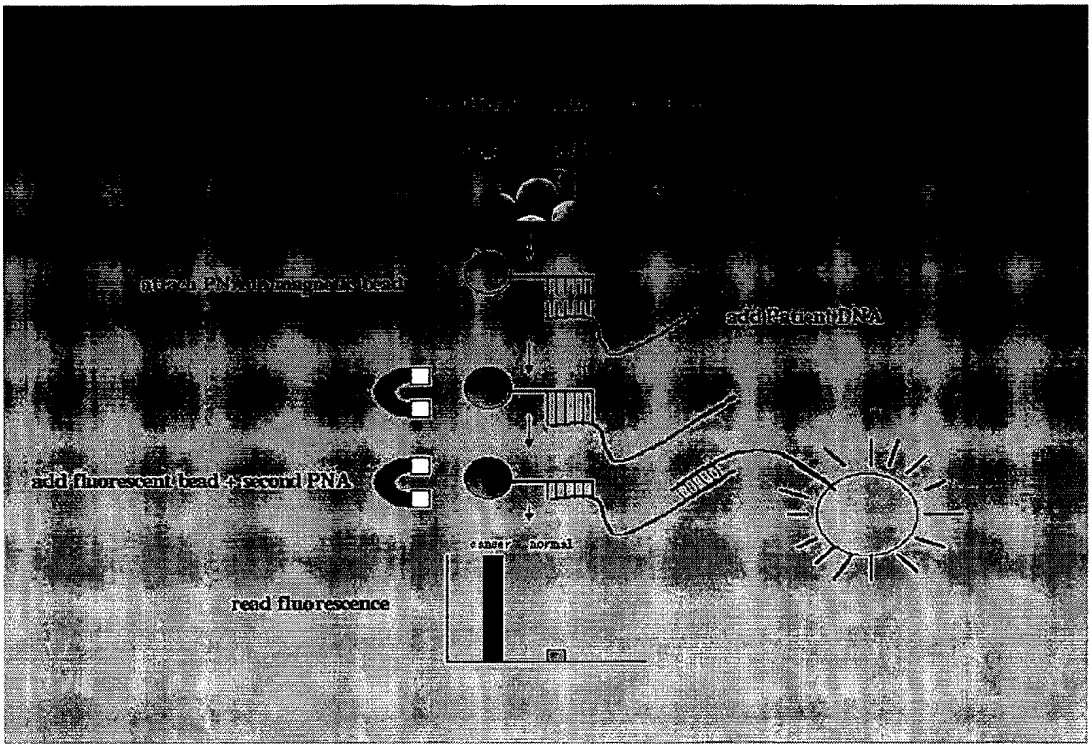


Figure 3

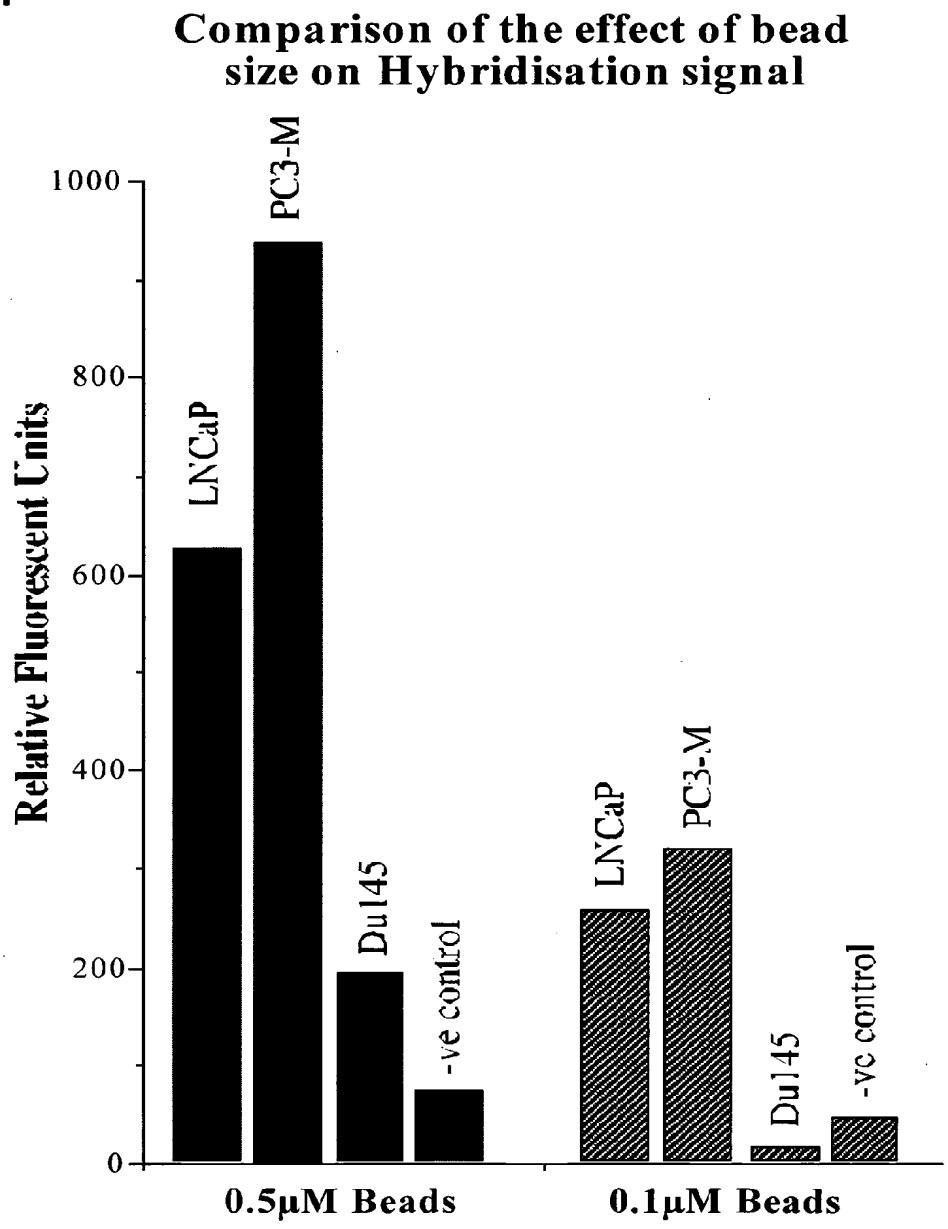
```
-39                    -38-37     -36-35-34     -33   -32
CAGCGGTCTT AGGGAATTTC CCCCCGCGAT GTCCCGGCGC GCCAGTTCGC TGCGCACACT -275
UAGUGGTUTT AGGGAATTTU UUUUUGTGAT GTUUUGGUGU GUUAGTTUGT TGUGUAUAUT  B-U
UAGCGGTUTT AGGGAATTTU UUUUCGCGAT GTUUCGGCGC GUUAGTTCGT TGCGUAUAUT  B-M


-31  -30                                        -29
TCGCTGCGGT CCTCTTCCTG CTGTCTGTTT ACTCCCTAGG CCCCGCTGGG GACCTGGGAA -215
TUGUTGUGGT UUTUTTUUTG UTGTUTGTTT AUTUUUTAGG UUUUGTTGGG GAUUTGGGAA B-U
TCGUTGCGGT UUTUTTUUTG UTGTUTGTTT AUTUUUTAGG UUUCGTTGGG GAUUTGGGAA B-M


                  -28       -27-26-25   -24             -23
AGAGGGAAAG GCTTCCCCGG CCAGCTGCGC GGCGACTCCG GGGACTCCAG GGCGCCCCTC -155
AGAGGGAAAG GUTTUUUUGG UUAGUTGUGU GGUGAUTUUG GGGAUTUUAG GGUGUUUUTU B-U
AGAGGGAAAG GUTTUUUCGG UUAGUTGCGC GGCGAUTUCG GGGAUTUUAG GGCGUUUUTU B-M
```

	Target
PNA#1 OEE- GAA ACA TCG CGA A -EE	Methylated DNA
PNA#2 OEE - GAA ACA TCG CGA AAA - EE	Methylated DNA
PNA#3 OEE - ATC GCC GCG CAA CTA A -EE	Methylated DNA
PNA#4 OEE - AAA ACA TCA CAA AAA - EE	Unethylated DNA
PNA#5 OEE - ATC ACC ACA CAA CTA A - EE	Unmethylated DNA
PNA#6 OEE - CTA ACG CGC CGA AAC - EE	Methylated DNA
PNA#7 OEE - CCA CTA CAA TCC CA - EE	Unmethylated DNA
PNA#8 OEE - CAC CAC ACA ACT - EE	Unmethylated DNA
PNA#9 OEE - GCA ACT AAG CAA CG -EE	Methylated DNA
PNA#10 Biotin-OEE - GCA ACG AAC TAA CG - EE	Methylated DNA

Figure 4



Methylated cell lines bisulphite modified LNCap and PC-3M

Unmethylated cell line bisulphite modified Du145

Controlunmodified ssDNA

Figure 5

Detection of prostate cancer cell lines and tissue DNA extracts using PNA technology

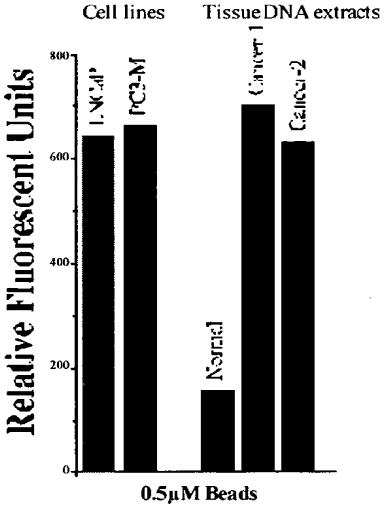


Figure 6

Effect of PNA concentration on the microtitre wells on sensitivity

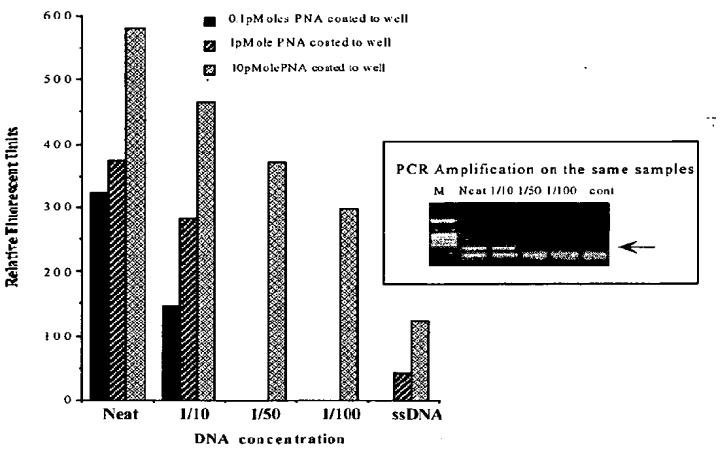


Figure 7

Single PNA methylation analysis using Oligreen™ detection agent

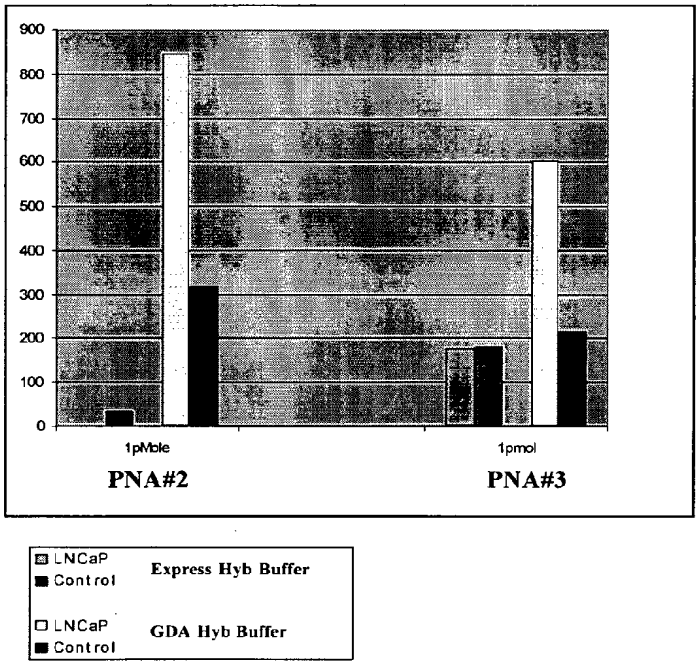


Figure 8

Detection of methylated molecules in a background of unmethylated sequences

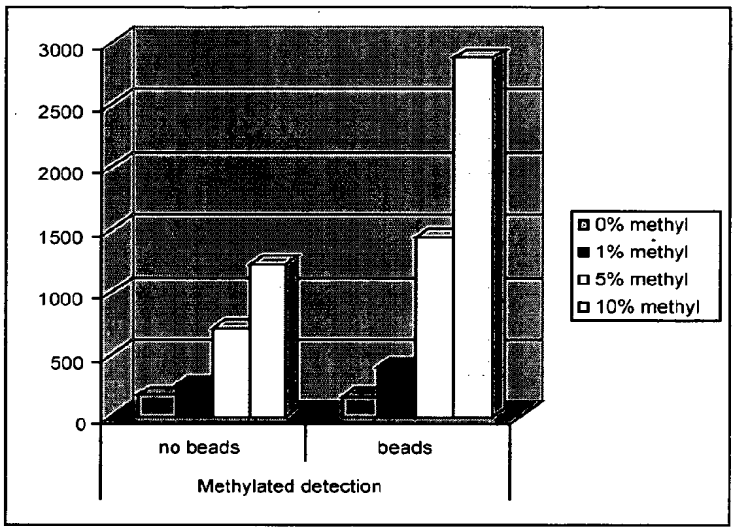


Figure 9

Detection of unmethylated molecules in a background of methylated sequences

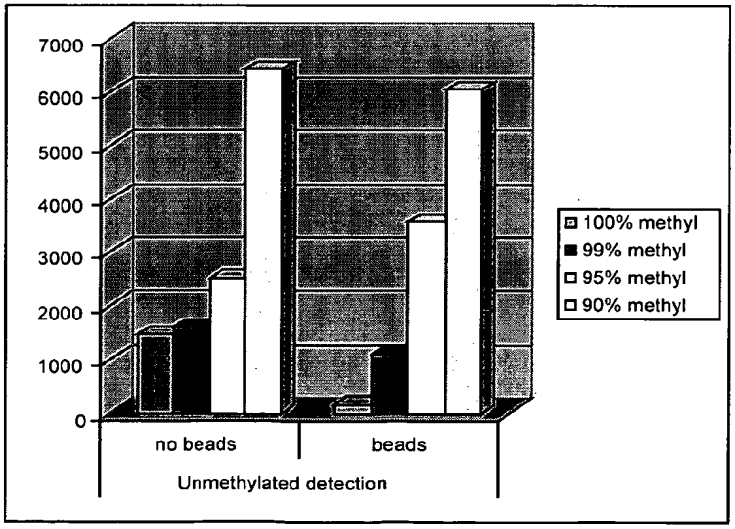
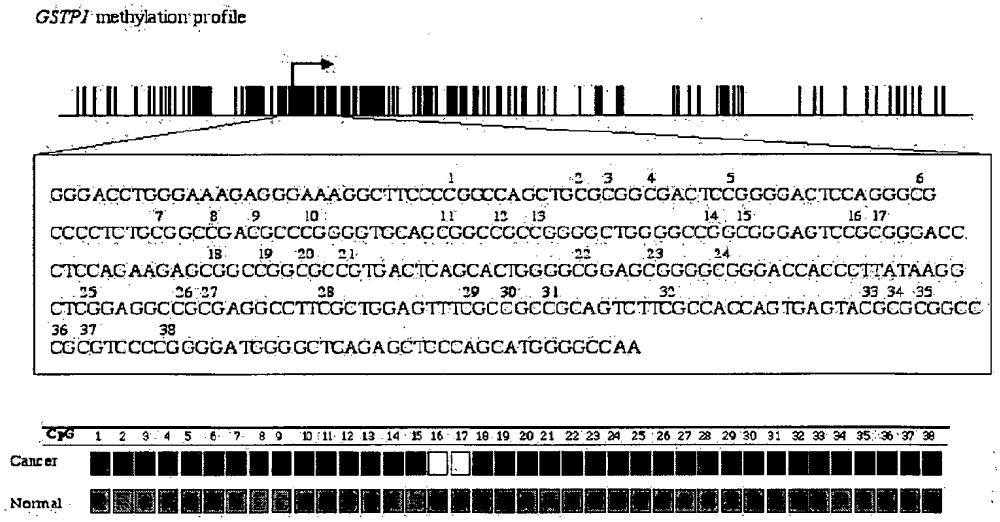


Figure 10


GSTP1 methylation profile
GGGACCTGGGAAAGAGGGAAAGGCTTCCCCGGCCAGCTGCGCGGCGACTCCGGGGACTCCAGGGCG
CCCCTCTGCGGCCGACGCCCGGGGTGCAGCGGCCGCCGGGGCTGGGGCCGGCGGGAGTCCGCGGGACC
CTCCAGAAGAGCGGCCGGCGCCGTGACTCAGCACTGGGGCGGAGCGGGGCGGGACCACCCTTATAAGG
CTCGGAGGCCGCGAGGCCTTCGCTGGAGTTTCGCCGCCGCAGTCTTCGCCACCAGTGAGTACGCGCGGCC
CGCGTCCCCGGGGATGGGGCTCAGAGCTCCCAGCATGGGGCCAA
CpG
Cancer
Normal

SEQUENCE LISTING

<110> Human Genetic Signatures Pty Ltd

<120> Detection of Nucleic Acid Molecules

<130> 204686143

<150> PR1425

<151> 2000-11-13

<160> 11

<170> PatentIn version 3.1

<210> 1
<211> 180
<212> DNA
<213> Homo sapiens

<400> 1

cagcggtctt agggaatttc cccccgcgat gtcccggcgc gccagttcgc tgcgcacact 60

tcgctgcggt cctcttcctg ctgtctgttt actccctagg ccccgctggg gacctgggaa 120

agagggaaag gcttccccgg ccagctgcgc ggcgactccg gggactccag ggcgcccctc 180

<210> 2

<211> 13

<212> DNA

<213> Homo sapiens

<400> 2

gaaacatcgc gaa 13

<210> 3

<211> 15

<212> DNA

<213> Homo sapiens

<400> 3

gaaacatcgc gaaaa 15

<210> 4

<211> 16

<212> DNA

<213> Homo sapiens

<400> 4
atcgccgcgc aactaa 16

<210> 5
<211> 15
<212> DNA
<213> Homo sapiens

<400> 5
aaaacatcac aaaaa 15

<210> 6
<211> 16
<212> DNA
<213> Homo sapiens

<400> 6
atcaccacac aactaa 16

<210> 7
<211> 15
<212> DNA
<213> Homo sapiens

<400> 7
ctaacgcgcc gaaac 15

<210> 8

<211> 14

<212> DNA

<213> Homo sapiens

<400> 8

ccactacaat ccca 14

<210> 9

<211> 12

<212> DNA

<213> Homo sapiens

<400> 9

caccacacaa ct 12

<210> 10

<211> 14

<212> DNA

<213> Homo sapiens

<400> 10

gcaactaagc aacg 14

<210> 11

<211> 14

<212> DNA

<213> Homo sapiens

<400> 11

gcaacgaact aacg 14

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 January 2017 (26.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/012544 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/CN2016/090642

(22) International Filing Date:
20 July 2016 (20.07.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/194,702 20 July 2015 (20.07.2015) US

(71) Applicant: **THE CHINESE UNIVERSITY OF HONG KONG** [CN/CN]; Office of Research and Knowledge Transfer Services, Room 301, Pi Ch'iu Building, Shatin, New Territories, Hong Kong 999077 (CN).

(72) Inventors: **LO, Yuk-Ming Dennis**; 4th Floor, 7 King Tak Street, Homantim, Kowloon, Hong Kong 999077 (CN). **CHAN, Kwan Chee**; Flat C, 15th Floor, Block 5, Peak One, Mei Tin Road, Shatin, New Territories, Hong Kong 999077 (CN). **CHIU, Rossa Wai Kwun**; House 31, Double Haven, 52 Ma Lok Path, Shatin, New Territories, Hong Kong 999077 (CN). **JIANG, Peiyong**; Flat 7 on 1st Floor of Block B of Kwong Lam Court, Nos. 62-66 (No.62) Siu Lek Yuen Road, Shatin, New Territories, Hong Kong 999077 (CN). **SUN, Kun**; Rm 711, Block A, Staff Quarters, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, New Territories, Hong Kong (CN).

(74) Agent: **INSIGHT INTELLECTUAL PROPERTY LIMITED**; 19 A, Tower A, InDo Building, No. 48A Zhichun Road, Haidian District, Beijing 100098 (CN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

WO 2017/012544 A1

(54) Title: METHYLATION PATTERN ANALYSIS OF HAPLOTYPES IN TISSUES IN DNA MIXTURE

(57) Abstract: Systems, apparatuses, and method are provided for determining the contributions of different tissues to a biological sample that includes a mixture of cell-free DNA molecules from various tissues types, e.g., as occurs in plasma or serum and other body fluids. Embodiments can analyze the methylation patterns of the DNA mixture (e.g., methylation levels at particular loci) for a particular haplotype and determine fractional contributions of various tissue types to the DNA mixture, e.g., of fetal tissue types or tissue types of specific organs that might have a tumor. Such fractional contributions determined for a haplotype can be used in a variety of ways.

METHYLATION PATTERN ANALYSIS OF HAPLOTYPES IN TISSUES IN DNA MIXTURE

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Application No. 62/194,702, entitled "Methylation Pattern Analysis Of Haplotypes In Tissues In A DNA Mixture" filed July 20, 2015, the entire contents of which is herein incorporated by reference for all purposes.

BACKGROUND

[0002] It has previously been demonstrated that through the analysis of plasma DNA of a pregnant woman carrying a fetus, the maternal haplotypes inherited by the fetus can be deduced using the process of relative haplotype dosage analysis (RHDO) (Lo et al. Sci Transl Med 2010; 2: 61ra91 and US Patent 8,467,976). The haplotype information for the pregnant woman can be used. The haplotype information can be obtained using family analysis or a method for the direct analysis of the haplotype (e.g. Fan et al. Nat Biotechnol 2011; 29: 51-57; Snyder et al. Nat Rev Genet 2015; 16: 344-358). SNPs that are heterozygous in the mother but homozygous in the father can be used for the RHDO analysis.

[0003] Such a use of specific SNPs can limit the loci that can be used, and therefore limit the amount of data and accuracy. Such use of specific SNPs may also limit the clinical utility of the method as DNA samples from additional family members may not be available, and methods for the direct analysis of haplotypes would add costs to the analysis.

BRIEF SUMMARY

[0004] Embodiments are described for determining the contributions of different tissues to a biological sample that includes a mixture of cell-free DNA molecules from various tissues types, e.g., as occurs in plasma or serum and other body fluids. Embodiments can analyze the methylation patterns of the DNA mixture (e.g., methylation levels at particular loci) for a particular haplotype and determine fractional contributions of various tissue types to the DNA

mixture, e.g., of fetal tissue types or tissue types of specific organs that might have a tumor. Such fractional contributions determined for a haplotype can be used in a variety of ways.

[0005] In some embodiments, two fractional contributions of a tissue type can be determined using methylation levels of two sets of cell-free DNA molecules from a maternal sample, each set being for a different one of two parental haplotypes of a parent of a fetus, for a chromosomal region being analyzed. In various implementations, the maternal sample can be the plasma or serum sample from a woman pregnant with one or more fetuses. The two fractional contributions can be used to identify a portion of the fetal genome. For example, a separation value between the two fractional contributions of fetal tissue can indicate the fetal genotype at a locus and can indicate which of the two parental haplotypes is inherited by the fetus. For instance, the higher fractional contribution can indicate the inherited haplotype, and both can be inherited if the separation value is less than a threshold; both haplotypes could be inherited when both parents share the haplotype (or allele for a genotype) for the region being analyzed.

[0006] In some embodiments, only one fractional contribution of fetal tissue is determined for one haplotype. When the one fractional contribution exceeds a reference value (e.g., as determined from other samples), the fetus can be determined to have inherited the one haplotype for the region being analyzed.

[0007] In some embodiments, two methylation levels can be determined for two sets of cell-free DNA molecules from a maternal sample, each set being for a different one of two parental haplotypes of a parent of a fetus, as part of identifying a portion of the fetal genome. The two methylation levels can be compared to each other to identify which haplotype is inherited by the fetus, e.g., by which methylation level is lower. For example, a fetus contributes cell-free DNA molecules that are hypomethylated, and a measurement of a lower methylation level of one haplotype indicates that the one haplotype is inherited by the fetus.

[0008] In some embodiments, a sequence imbalance can be detected for a target chromosomal region of a fetus using a mixture of cell-free DNA molecules from a plurality of tissues types. Target heterozygous loci can be identified for the target chromosomal region having a first target haplotype and a second target haplotype having different alleles. A first target fractional contribution of the fetal tissue type in the mixture can be determined using methylation levels at the target heterozygous loci, where the methylation levels are determined using a target set of

cell-free DNA molecules located at (i.e., covering) the loci of the first haplotype. Similarly, a first reference fractional contribution of the fetal tissue type can be determined. A separation value of the first target fractional contribution and the first reference fractional contribution can be compared to a threshold value to determine whether the fetus has a sequence imbalance. If the two fractional contributions are significantly different, then a sequence imbalance can be determined. The specific threshold being used can depend on the specific sequence imbalance (e.g., an amplification or a deletion) being tested.

[0009] In some embodiments, a fractional contribution of a first haplotype in a first tissue type can be used to determine whether the first tissue type has a disease state. The first haplotype can have a signature specific to healthy cells or to abnormal cells. Thus, the first haplotype can be not present in healthy cells of the organism, or present in healthy cells of the organism and not present in abnormal cells that may be in the mixture. A separation value between the first fractional contribution and a reference fractional contribution can be compared to a threshold value to determine a classification of whether the first tissue type has a disease state.

[0010] In some embodiments, the tissue origin of a copy number aberration can be determined using methylation deconvolution. A first chromosomal region can be identified as exhibiting a copy number aberration. The haplotypes in the first chromosomal region can be determined. For each of M tissue types, a corresponding separation value between the corresponding first fractional contribution and the corresponding second fractional contribution can be determined. The tissue type with the highest separation value can be identified as the tissue of origin.

[0011] Other embodiments are directed to systems and computer readable media associated with methods described herein.

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

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 is a flowchart illustrating a method of analyzing a DNA mixture of cell-free DNA molecules to determine fractional contributions from various tissue types from methylation levels according to embodiments of the present invention.

[0014] FIG. 2 shows a schematic diagram showing several potential applications of DNA methylation deconvolution (e.g., using plasma) and its applications according to embodiments of the present invention.

[0015] FIG. 3A shows a graph of percentage contributions of different organs to plasma DNA for 15 pregnant women according to embodiments of the present invention. FIG. 3B shows a plot 350 of a correlation between the plasma DNA fractions contributed by the placenta deduced from plasma DNA methylation deconvolution and the fetal DNA fractions deduced using fetal-specific SNP alleles according to embodiments of the present invention.

[0016] FIG. 4 shows a table of percentage contributions determined from a plasma DNA tissue mapping analysis among pregnant women according to embodiments of the present invention.

[0017] FIG. 5 shows plots of percentage contributions of organs other than the placenta by plasma DNA tissue mapping and fetal DNA fractions based on fetal-specific SNP alleles according to embodiments of the present invention.

[0018] FIG. 6 shows a table of percentage contributions from plasma DNA tissue mapping analysis among the non-pregnant healthy control subjects according to embodiments of the present invention.

[0019] FIG. 7 shows a table of the estimated contributions of different organs to the plasma DNA for 11 pregnant women and 4 non-pregnant healthy subjects using the first set of markers (with high organ specificity) according to embodiments of the present invention.

[0020] FIG. 8 shows a table of the estimated contributions of different organs to the plasma DNA for 11 pregnant women and 4 non-pregnant healthy subjects using the second set of markers (with low organ specificity) according to embodiments of the present invention.

[0021] FIG. 9A is a plot showing the correlation between the estimated fetal DNA fraction (contribution from the placenta) and the fetal DNA fraction determined by counting the fetal-specific alleles in the maternal plasma samples.

[0022] FIG. 9B is a plot showing absolute difference between the estimation from methylation markers and fetal DNA fraction determined by fetal-specific alleles counting.

[0023] FIG. 10A is a graph showing placental contribution to plasma DNA deduced using markers with different selection criteria according to embodiments of the present invention. FIG. 10B is a graph showing the accuracy of plasma DNA deconvolution using markers with low variability (category i) and high variability (category ii) in the same type of tissue.

[0024] FIG. 11A shows a first scenario where the fetus has inherited the M allele from the mother and has a genotype of MN at a particular locus according to embodiments of the present invention. FIG. 11B shows a second scenario where the fetus has inherited the N allele from the mother and has the genotype of NN at a particular locus according to embodiments of the present invention.

[0025] FIG. 12A shows a determination of a maternal haplotype inherited by a fetus using methylation deconvolution according to embodiments of the present invention. FIG. 12B shows an illustration of the paternal haplotype methylation analysis according to embodiments of the present invention.

[0026] FIG. 13 is a flowchart illustrating a method 1300 of determining a portion of a fetal genome from a maternal sample using methylation deconvolution according to embodiments of the present invention.

[0027] FIG. 14 is a flowchart illustrating a method 1400 of determining a portion of a fetal genome from a maternal sample using methylation levels according to embodiments of the present invention.

[0028] FIG. 15 shows chromosomal aneuploidy detection based on haplotype deconvolution for maternal haplotypes according to embodiments of the present invention.

[0029] FIG. 16 shows chromosomal aneuploidy detection based on haplotype deconvolution for paternal haplotypes according to embodiments of the present invention.

[0030] FIG. 17 is a flowchart of a method 1700 for detecting a sequence imbalance in a portion of a fetal genome of an unborn fetus of a pregnant female using a biological sample from the pregnant female according to embodiments of the present invention.

[0031] FIG. 18 shows an illustration of haplotype deconvolution for organ transplant monitoring according to embodiments of the present invention.

[0032] FIG. 19 is a flowchart illustrating a method of analyzing a biological sample of an organism to detect whether a first tissue type has a disease state associated with a first haplotype according to embodiments of the present invention.

[0033] FIG. 20 shows a plot of copy number aberrations detected in the plasma of an HCC patient according to embodiments of the present invention.

[0034] FIG. 21 is a flowchart illustrating a method of analyzing a biological sample of an organism to identify an origin of a chromosomal aberration according to embodiments of the present invention.

[0035] FIG. 22 shows a block diagram of an example computer system 10 usable with system and methods according to embodiments of the present invention.

TERMS

[0036] A "methylome" provides a measure of an amount of DNA methylation at a plurality of sites or loci in a genome. The methylome may correspond to all of the genome, a substantial part of the genome, or relatively small portion(s) of the genome. Examples of methylomes of interest are the methylomes of organs (e.g. methylomes of brain cells, bones, the lungs, the heart, the muscles and the kidneys, etc.) that can contribute DNA into a bodily fluid (e.g. plasma, serum, sweat, saliva, urine, genital secretions, semen, stools fluid, diarrheal fluid, cerebrospinal fluid, secretions of the gastrointestinal tract, ascitic fluid, pleural fluid, intraocular fluid, fluid from a hydrocele (e.g. of the testis), fluid from a cyst, pancreatic secretions, intestinal secretions, sputum, tears, aspiration fluids from breast and thyroid, etc.). The organs may be transplanted organs. The methylome of a fetus is another example.

[0037] A "plasma methylome" is a methylome determined from the plasma or serum of an animal (e.g., a human). The plasma methylome is an example of a cell-free methylome since

plasma and serum include cell-free DNA. The plasma methylome is also an example of a mixed methylome since it is a mixture of fetal/maternal methylome or tumor/patient methylome or DNA derived from different tissues or organs or donor/recipient methylome in the context or organ transplantation.

[0038] A "site" (also called a "genomic site") corresponds to a single site, which may be a single base position or a group of correlated base positions, e.g., a CpG site or larger group of correlated base positions. A "locus" may correspond to a region that includes multiple sites. A locus can include just one site, which would make the locus equivalent to a site in that context.

[0039] The "methylation index" for each genomic site (e.g., a CpG site) can refer to the proportion of DNA fragments (e.g., as determined from sequence reads or probes) showing methylation at the site over the total number of reads covering that site. A "read" can correspond to information (e.g., methylation status at a site) obtained from a DNA fragment. A read can be obtained using reagents (e.g. primers or probes) that preferentially hybridize to DNA fragments of a particular methylation status. Typically, such reagents are applied after treatment with a process that differentially modifies or differentially recognizes DNA molecules depending of their methylation status, e.g. bisulfite conversion, or methylation-sensitive restriction enzyme, or methylation binding proteins, or anti-methylcytosine antibodies, or single molecule sequencing techniques that recognize methylcytosines and hydroxymethylcytosines.

[0040] The "methylation density" of a region can refer to the number of reads at sites within the region showing methylation divided by the total number of reads covering the sites in the region. The sites may have specific characteristics, e.g., being CpG sites. Thus, the "CpG methylation density" of a region can refer to the number of reads showing CpG methylation divided by the total number of reads covering CpG sites in the region (e.g., a particular CpG site, CpG sites within a CpG island, or a larger region). For example, the methylation density for each 100-kb bin in the human genome can be determined from the total number of cytosines not converted after bisulfite treatment (which corresponds to methylated cytosine) at CpG sites as a proportion of all CpG sites covered by sequence reads mapped to the 100-kb region. This analysis can also be performed for other bin sizes, e.g. 500 bp, 5 kb, 10 kb, 50-kb or 1-Mb, etc. A region could be the entire genome or a chromosome or part of a chromosome (e.g. a chromosomal arm). The methylation index of a CpG site is the same as the methylation density for a region when the

region only includes that CpG site. The "proportion of methylated cytosines" can refer the number of cytosine sites, "C's", that are shown to be methylated (for example unconverted after bisulfite conversion) over the total number of analyzed cytosine residues, i.e. including cytosines outside of the CpG context, in the region. The methylation index, methylation density and proportion of methylated cytosines are examples of "methylation levels." Apart from bisulfite conversion, other processes known to those skilled in the art can be used to interrogate the methylation status of DNA molecules, including, but not limited to enzymes sensitive to the methylation status (e.g. methylation-sensitive restriction enzymes), methylation binding proteins, single molecule sequencing using a platform sensitive to the methylation status (e.g. nanopore sequencing (Schreiber et al. Proc Natl Acad Sci 2013; 110: 18910-18915) and by the Pacific Biosciences single molecule real time analysis (Flusberg et al. Nat Methods 2010; 7: 461-465)).

[0041] A "methylation profile" (also called methylation status) includes information related to DNA methylation for a region. Information related to DNA methylation can include, but not limited to, a methylation index of a CpG site, a methylation density of CpG sites in a region, a distribution of CpG sites over a contiguous region, a pattern or level of methylation for each individual CpG site within a region that contains more than one CpG site, and non-CpG methylation. A methylation profile of a substantial part of the genome can be considered equivalent to the methylome. "DNA methylation" in mammalian genomes typically refers to the addition of a methyl group to the 5' carbon of cytosine residues (i.e. 5-methylcytosines) among CpG dinucleotides. DNA methylation may occur in cytosines in other contexts, for example CHG and CHH, where H is adenine, cytosine or thymine. Cytosine methylation may also be in the form of 5-hydroxymethylcytosine. Non-cytosine methylation, such as N^6-methyladenine, has also been reported.

[0042] "Methylation-aware sequencing" refers to any sequencing method that allows one to ascertain the methylation status of a DNA molecule during a sequencing process, including, but not limited to bisulfite sequencing, or sequencing preceded by methylation-sensitive restriction enzyme digestion, immunoprecipitation using anti-methylcytosine antibody or methylation binding protein, or single molecule sequencing that allows elucidation of the methylation status.

[0043] A "tissue" corresponds to a group of cells that group together as a functional unit. More than one type of cells can be found in a single tissue. Different types of tissue may consist of

different types of cells (e.g., hepatocytes, alveolar cells or blood cells), but also may correspond to tissue from different organisms (mother vs. fetus) or to healthy cells vs. tumor cells. "Reference tissues" correspond to tissues used to determine tissue-specific methylation levels. Multiple samples of a same tissue type from different individuals may be used to determine a tissue-specific methylation level for that tissue type.

[0044] A "biological sample" refers to any sample that is taken from a subject (*e.g.*, a human, such as a pregnant woman, a person with cancer, or a person suspected of having cancer, an organ transplant recipient or a subject suspected of having a disease process involving an organ (e.g., the heart in myocardial infarction, or the brain in stroke, or the hematopoietic system in anemia) and contains one or more nucleic acid molecule(s) of interest. The biological sample can be a bodily fluid, such as blood, plasma, serum, urine, vaginal fluid, fluid from a hydrocele (e.g. of the testis), or vaginal flushing fluids, pleural fluid, ascitic fluid, cerebrospinal fluid, saliva, sweat, tears, sputum, bronchoalveolar lavage fluid, etc. Stool samples can also be used. In various embodiments, the majority of DNA in a biological sample that has been enriched for cell-free DNA (e.g., a plasma sample obtained via a centrifugation protocol) can be cell-free (as opposed to cells), e.g., greater than 50%, 60%, 70%, 80%, 90%, 95%, or 99%. The centrifugation protocol can include 3,000 g x 10 minutes, obtaining the fluid part, and recentrifuging at 30,000 g for another 10 minutes to remove residual cells.

[0045] The term "level of cancer" can refer to whether cancer exists (i.e., presence or absence), a stage of a cancer, a size of tumor, whether there is metastasis, the total tumor burden of the body, and/or other measure of a severity of a cancer (e.g. recurrence of cancer). The level of cancer could be a number or other indicia, such as symbols, alphabet letters, and colors. The level could be zero. The level of cancer also includes premalignant or precancerous conditions (states) associated with mutations or a number of mutations. The level of cancer can be used in various ways. For example, screening can check if cancer is present in someone who is not known previously to have cancer. Assessment can investigate someone who has been diagnosed with cancer to monitor the progress of cancer over time, study the effectiveness of therapies or to determine the prognosis. In one embodiment, the prognosis can be expressed as the chance of a patient dying of cancer, or the chance of the cancer progressing after a specific duration or time, or

the chance of cancer metastasizing. Detection can mean 'screening' or can mean checking if someone, with suggestive features of cancer (e.g. symptoms or other positive tests), has cancer.

[0046] The term "sequence imbalance" of a chromosomal region can refer to any significant deviation in an amount of cell-free DNA molecules from the chromosomal region relative to an expected value, if the organism was healthy. For example, a chromosomal region may exhibit an amplification or a deletion in a certain tissue, thereby resulting in a sequence imbalance for the chromosomal region in a DNA mixture containing DNA from the tissue, mixed with DNA from other tissues. As examples, the expected value can be obtained from another sample or from another chromosomal region that is assumed to be normal (e.g., an amount representative of two copies for a diploid organism). A chromosomal region can be composed of multiple disjoint subregions.

[0047] A "type" for a genomic locus (marker) corresponds to specific attributes for a locus across tissue types. The description primarily refers to type I loci and type II loci, whose properties are provided in detail below. A locus of a given type can have specific statistical variation in methylation levels across tissue types. A "category" for a genomic locus (marker) corresponds to specific variation in methylation levels for a locus across different individuals for a same tissue type. A set of genomic loci (markers) can be composed of any number of loci of various types and/or categories. Thus, a set of loci corresponds to loci selected for a particular measurement and does not connote any particular properties of the loci in the set.

[0048] A "separation value" corresponds to a difference or a ratio involving two values, e.g., two fractional contributions or two methylation levels. The separation value could be a simple difference or ratio. The separation value can include other factors, e.g., multiplicative factors. As other examples, a difference or ratio of functions of the values can be used, e.g., a difference or ratio of the natural logarithms (ln) of the two values. A separation value can include a difference and a ratio.

[0049] The term "*classification*" as used herein refers to any number(s) or other characters(s) that are associated with a particular property of a sample. For example, a "+" symbol (or the word "positive") could signify that a sample is classified as having deletions or amplifications. The classification can be binary (e.g., positive or negative) or have more levels of classification (e.g., a scale from 1 to 10 or 0 to 1). The term "*cutoff*" and "*threshold*" refer to a predetermined number

used in an operation. A threshold value may be a value above or below which a particular classification applies. Either of these terms can be used in either of these contexts.

DETAILED DESCRIPTION

[0050] Methylation differences among tissue types (e.g., fetal tissues, liver etc.) in a DNA mixture (e.g., plasma) may be used to differentiate properties of haplotypes in a specific tissue type. For example, the methylation levels of two maternal haplotypes in the plasma of a pregnant woman can be used to determine which haplotype is inherited from the mother to the fetus. As another example, the methylation levels of two haplotypes in fetal tissue can be used to detect a sequence imbalance (e.g., an aneuploidy) in the fetus. Other tissue types may also be analyzed, e.g., to detect a disease state in a specific tissue type. The tissue type from which a copy number aberration originates may also be determined.

[0051] Some embodiments can determine percentages of cell-free DNA in plasma (or other DNA mixture) from various tissue types using known methylation levels at certain genomic sites for the specific tissue types. For example, methylation levels at the genomic sites can be measured for a liver sample, and these tissue-specific methylation levels can be used to determine how much cell-free DNA in the mixture is from the liver. Methylation levels can be measured for tissue types that provide substantial contributions to the DNA mixture, so that a predominance (e.g., more than 90%, 95%, or 99%) of the cell-free DNA mixture can be accounted for. Such other samples can include, but not limited to, some or all of the following: lung, colon, small intestines, pancreas, adrenal glands, esophagus, adipose tissues, heart, and brain.

[0052] A deconvolution process can be used to determine fractional contributions (e.g., percentage) for each of the tissue types for which tissue-specific methylation levels are known. In some embodiments, a linear system of equations can be created from the known tissue-specific methylation levels and the mixture methylation levels at the specified genomic sites, and the fractional contributions that best approximate the measured mixture methylation levels can be determined (e.g., using least squares).

[0053] Once the fractional contributions are determined, the fractional contributions can be used for various purposes. For example, differences in fractional contributions of fetal tissue can be used to determine which haplotype is inherited from a parent. The alleles at one or more

heterozygous loci can be determined for each of two parental haplotypes. Cell-free DNA at the one or more heterozygous loci can be used to determine two fractional contributions: one for each haplotype. For example, cell-free DNA molecules having alleles of a first haplotype can be used to determine a first fractional contribution, and cell-free DNA molecules having alleles of a second haplotype can be used to determine a second fractional contribution. The inherited haplotype will correspond to the higher fractional contribution for fetal tissue.

[0054] Further, an inherited haplotype will have lower methylation levels due to the general hypomethylation of fetal cell-free DNA. The methylation levels for the two haplotypes can be compared, and the haplotype with lower methylation level can be identified as being the inherited haplotype.

[0055] As another example, a sequence imbalance can be detected in a target chromosomal region of a fetus. A target fractional contribution of the fetal tissue type in the mixture can be determined for a first haplotype in the target chromosomal region. Similarly, a reference fractional contribution of the fetal tissue type can be determined for a reference chromosomal region. A separation value between the two contributions can be compared to a threshold value to determine whether the fetus has a sequence imbalance (e.g., an aneuploidy).

[0056] As another example, a first haplotype can have a signature specific to healthy cells or to abnormal cells. A separation value between the fractional contribution determined for the first haplotype and a reference fractional contribution can be compared to a threshold value to determine a classification of whether the first tissue type has a disease state. As examples, the first haplotype can be of a transplanted organ or a tumor, or only be in healthy cells and not in a transplanted organ or a tumor. The disease state may be whether the transplanted organ is being rejected, or whether a tumor is increasing in size or metastasized (e.g., after a surgery did not remove all of the tumor).

[0057] As another example, the tissue origin of a copy number aberration can be determined using methylation deconvolution. A first chromosomal region can be identified as exhibiting a copy number aberration. For each of M tissue types, a corresponding separation value between the fractional contributions of the two haplotypes in the first chromosomal region can be determined. The tissue type with the highest separation value can be identified as the tissue of origin.

[0058] Methylation deconvolution is first described, and then the selection of methylation markers and the accuracy of the methylation deconvolution are described. The use of the fractional contributions to determine part of a fetal genome is then described.

I. COMPOSITION OF DNA MIXTURE BY METHYLATION DECONVOLUTION

[0059] Different tissue types can have different levels of methylation for a genomic site. These differences can be used to determine the fractional contributions of DNA from the various tissue types in a mixture. Thus, the composition of a DNA mixture can be determined by a tissue-specific methylation pattern analysis. The examples below discuss methylation densities, but other methylation levels can be used.

A. Single Genomic Site

[0060] The principle of methylation deconvolution can be illustrated using a single methylation genomic site (methylation marker) to determine a composition of a DNA mixture from an organism. Assume that tissue A is completely methylated for the genomic site, i.e. methylation density (MD) of 100% and tissue B is completely unmethylated, i.e. MD of 0%. In this example, methylation density refers to the percentage of cytosine residues with the context of CpG dinucleotides being methylated in the region of interest.

[0061] If the DNA mixture C is composed of tissue A and tissue B and the overall methylation density of the DNA mixture C is 60%, we can deduce the proportional contribution of tissues A and B to the DNA mixture C according to the following formula:

$$MD_C = MD_A \times a + MD_B \times b,$$

where MD_A, MD_B, MD_C represent the MD of tissues A, tissue B and the DNA mixture C, respectively; and a and b are the proportional contributions of tissues A and B to the DNA mixture C. In this particular example, it is assumed that tissues A and B are the only two constituents of the DNA mixture. Therefore, a + b = 100%. Thus, it is calculated that tissues A and B contribute 60% and 40%, respectively, to the DNA mixture.

[0062] The methylation densities in tissue A and tissue B can be obtained from samples of the organism or from samples from other organisms of the same type (e.g., other humans, potentially

of a same subpopulation). If samples from other organisms are used, a statistical analysis (e.g., average, median, geometric mean) of the methylation densities of the samples of tissue A can be used to obtain the methylation density MD_A, and similarly for MD_B.

[0063] Genomic site can be chosen to have minimal inter-individual variation, for example, less than a specific absolute amount of variation or being within a lowest portion of genomic sites tested. For instance, for the lowest portion, embodiments can select only genomic sites having the lowest 10% of variation among a group of genomic sites tested. The other organisms can be taken from healthy persons, as well as those with particular physiologic conditions (e.g. pregnant women, or people with different ages or people of a particular sex), which may correspond to a particular subpopulation that includes the current organism being tested.

[0064] The other organisms of a subpopulation may also have other pathologic conditions (e.g. patients with hepatitis or diabetes, etc.). Such a subpopulation may have altered tissue-specific methylation patterns for various tissues. The methylation pattern of the tissue under such disease condition can be used for the deconvolution analysis in addition to using the methylation pattern of the normal tissue. This deconvolution analysis may be more accurate when testing an organism from such a subpopulation with those conditions. For example, a cirrhotic liver or a fibrotic kidney may have a different methylation pattern compared with a normal liver and normal kidney, respectively. Thus, if a patient with liver cirrhosis was screened for other diseases, it can be more accurate to include a cirrhotic liver as one of the candidates contributing DNA to the plasma DNA, together with the healthy tissues of other tissue types.

B. Multiple Genomic Sites

[0065] More genomic sites (e.g., 10 or more) may be used to determine the constitution of the DNA mixture when there are more potential candidate tissues. The accuracy of the estimation of the proportional composition of the DNA mixture is dependent on a number of factors including the number of genomic sites, the specificity of the genomic sites (also called "sites") to the specific tissues, and the variability of the sites across different candidate tissues and across different individuals used to determine the reference tissue-specific levels. The specificity of a site to a tissue refers to the difference in the methylation density of the genomic sites between the particular tissue and other tissue types.

[0066] The larger the difference between their methylation densities, the more specific the site to the particular tissue would be. For example, if a site is completely methylated in the liver (methylation density = 100%) and is completely unmethylated in all other tissues (methylation density = 0%), this site would be highly specific for the liver. Whereas, the variability of a site across different tissues can be reflected by, for example, but not limited to, the range or standard deviation of methylation densities of the site in different types of tissue. A larger range or higher standard deviation would allow a more precise and accurate determination of the relative contributions of the different organs to the DNA mixture mathematically. The effects of these factors on the accuracy of estimating the proportional contribution of the candidate tissues to the DNA mixture are illustrated in the later sections of this application.

[0067] Here, we use mathematical equations to illustrate the deduction of the proportional contribution of different organs to the DNA mixture. The mathematical relationship between the methylation densities of the different sites in the DNA mixture and the methylation densities of the corresponding sites in different tissues can be expressed as:

$$\overline{MD_i} = \sum_k (p_k \times MD_{ik}),$$

where $\overline{MD_i}$ represents the methylation density of the site i in the DNA mixture; p_k represents the proportional contribution of tissue k to the DNA mixture; MD_{ik} represents the methylation density of the site i in the tissue k. When the number of sites is the same or larger than the number of organs, the values of individual p_k can be determined. The tissue-specific methylation densities can be obtained from other individuals, and the sites can be chosen to have minimal inter-individual variation, as mentioned above.

[0068] Additional criteria can be included in the algorithm to improve the accuracy. For example, the aggregated contribution of all tissues can be constrained to be 100%, i.e.

$$\sum_k p_k = 100\%.$$

Furthermore, all the organs' contributions can be required to be non-negative:

$$p_k \geq 0, \forall k$$

[0069] Due to biological variations, the observed overall methylation pattern may not be completely identical to the methylation pattern deduced from the methylation of the tissues. In

such a circumstance, mathematical analysis would be required to determine the most likely proportional contribution of the individual tissues. In this regard, the difference between the observed methylation pattern in the DNA and the deduced methylation pattern from the tissues is denoted by W.

$$W = O - \sum_k (p_k \times M_k)$$

where O is the observed methylation pattern for the DNA mixture and M_k is the methylation pattern of the individual tissue k. p_k is the proportional contribution of tissue k to the DNA mixture. The most likely value of each p_k can be determined by minimizing W, which is the difference between the observed and deduced methylation patterns. This equation can be resolved using mathematical algorithms, for example by, but not limited to, using quadratic programming, linear/non-linear regression, expectation-maximization (EM) algorithm, maximum likelihood algorithm, maximum a posteriori estimation, and the least squares method.

C. Method of Methylation Deconvolution

[0070] As described above, a biological sample including a mixture of cell-free DNA molecules from an organism can be analyzed to determine the composition of the mixture, specifically the contributions from different tissue types. For example, the percentage contribution of the cell-free DNA molecules from the liver can be determined. These measurements of the percentage contributions in the biological sample can be used to make other measurements of the biological sample, e.g., identifications of where a tumor is located, as is described in later sections.

[0071] FIG. 1 is a flowchart illustrating a method 100 of analyzing a DNA mixture of cell-free DNA molecules to determine fractional contributions from various tissue types from methylation levels according to embodiments of the present invention. A biological sample includes a mixture of cell-free DNA molecules from M tissues types. The biological sample can be any one of various examples, e.g., as mentioned herein. The number M of tissue types is greater than two. In various embodiments, M can be 3, 7, 10, 20, or more, or any number in between. Method 100 can be performed at least partially using a computer system, as can other methods described herein.

[0072] At block 110, N genomic sites are identified for analysis. The N genomic sites can have various attributes, e.g., as described in more detail in section II, which describes type I and type II

genomic sites. As examples, the N genomic sites can include type I or type II sites only, or a combination of both. The genomic sites can be identified based on analyses of one or more other samples, e.g., based on data obtained from databases about methylation levels measured in various individuals.

[0073] Specific genomic sites can be selected to provide a desired level of accuracy. For example, loci that have at least a threshold variability can be used, as opposed to just using loci that are specific to one tissue type. A first set (e.g., 10) of the genomic sites can be selected such that each have a coefficient of variation of methylation levels of at least 0.15 across M tissue types and such that each has a difference between a maximum and a minimum methylation level for the M tissue types that exceeds 0.1 for one or more other samples. This first set of genomic sites may not have a specific methylation signature for a specific tissue type, e.g., only or predominantly methylated in the specific tissue type. Such a first set is referred to as type II sites. These genomic sites can be used in combination with genomic sites that do have a specific signature, which are referred to as type I sites.

[0074] Using the type II sites can ensure that the full space of methylation levels across the tissue types is spanned by the genomic sites, thereby providing increased accuracy over the type I sites. Just using more type I sites provides redundant basis vectors for the methylation space (i.e., more genomic sites that have the same pattern as other sites), while adding other genomic sites whose methylation levels have various values across different tissues adds new basis vectors for discriminating fractional contributions via the linear system of equations.

[0075] In some embodiments, at least 10 of the N genomic sites each have a coefficient of variation of methylation levels of at least 0.15 across the M tissue types. The at least 10 genomic sites can also each have a difference between a maximum and a minimum methylation level for the M tissue types that exceeds 0.1. These methylation properties of the genomic loci can be measured for one sample or a set of samples. The set of samples may be for a subpopulation of organisms that includes the instant organism being tested, e.g., a subpopulation having a particular trait that is shared with the instant organism. These other samples can be referred to as reference tissues, and different reference tissues may be used from different samples.

[0076] At block 120, N tissue-specific methylation levels are obtained at the N genomic sites for each of M tissue types. N is greater than or equal to M, so that the tissue-specific methylation

levels can be used in the deconvolution to determine the fractional percentages. The tissue-specific methylation levels can form a matrix A of dimensions N by M. Each column of the matrix A can correspond to a methylation pattern for a particular tissue type, where the pattern is of methylation levels at the N genomic sites.

[0077] In various embodiments, the tissue-specific methylation patterns can be retrieved from public database(s) or previous studies. In examples herein, the methylation data for neutrophils and B cells were downloaded from the Gene Expression Omnibus (Hodges et al. Mol Cell 2011;44:17-28). Methylation patterns for other tissues (hippocampus, liver, lung, pancreas, atrium, colon (including its various parts, e.g. sigmoid colon, transverse colon, ascending colon, descending colon), adrenal gland, esophagus, small intestines and CD4 T cell) were downloaded from the RoadMap Epigenomics project (Ziller et al. Nature 2013;500:477-81). The methylation patterns for the buffy coat, placenta, tumor and plasma data were from published reports (Lun et al. Clin Chem. 2013;59:1583-94; Chan et al. Proc Natl Acad Sci U S A. 2013;110:18761-8). These tissue-specific methylation patterns can be used to identify the N genomic sites to be used in the deconvolution analysis.

[0078] At block 130, the biological sample including a mixture of cell-free DNA molecules from the M tissues types is received. The biological sample may be obtained from the patient organism in a variety of ways. The manner of obtaining such samples may be non-invasive or invasive. Examples of non-invasively obtained samples include certain types of fluids (e.g. plasma or serum or urine) or stools. For instance, plasma includes cell-free DNA molecules from many organ tissues, and is thus useful for analyzing many organs via one sample.

[0079] At block 140, cell-free DNA molecules from the biological sample are analyzed to identify their locations in a reference genome corresponding to the organism. For example, the cell-free DNA molecules can be sequenced to obtain sequence reads, and the sequence reads can be mapped (aligned) to the reference genome. If the organism was a human, then the reference genome would be a reference human genome, potentially from a particular subpopulation. As another example, the cell-free DNA molecules can be analyzed with different probes (e.g., following PCR or other amplification), where each probe corresponds to a genomic location, which may cover a heterozygous and one or more CpG sites, as is described below.

[0080] A statistically significant number of cell-free DNA molecules can be analyzed so as to provide an accurate deconvolution for determining the fractional contributions from the M tissue types. In some embodiments, at least 1,000 cell-free DNA molecules are analyzed. In other embodiments, at least 10,000 or 50,000 or 100,000 or 500,000 or 1,000,000 or 5,000,000 cell-free DNA molecules or more can be analyzed. The total number of molecules to analyze can depend on M and N, and the desired precision (accuracy). In various examples, the total number of cell-free DNA analyzes can be less than 500,000, one million, two million, five million, ten millsion, 20 million, or 50 million.

[0081] At block 150, N mixture methylation levels are measured at the N genomic sites using cell-free DNA molecules that are each located at any one of N genomic sites of the reference genome. A DNA molecule can be identified as located at a genomic site or a locus by one or more bases of the DNA molecule corresponding to one or more base positions of the genomic site or locus. Thus, the sequence of the DNA molecule would cover one or more base positions of the genomic site or locus. This information can be determined based on the locations determined in block 140. Such an identification of a DNA molecule located at a site of a locus can be used for any similar block of methods described herein.

[0082] The N mixture methylation levels refer to methylation levels in the mixture of the biological sample. As an example, if a cell-free DNA molecule from the mixture is located at one of the N genomic sites, then a methylation index for that molecule at the site can be included in an overall methylation density for that site. The N mixture methylation levels can form a methylation vector b of length N, where b corresponds to observed values from which the fractional contribution of each corresponding tissue type can be determined.

[0083] In one embodiment, the methylation levels for the genomic sites in the DNA mixture can be determined using whole genome bisulfite sequencing. In other embodiments, the methylation levels for the CpG sites can be determined using methylation microarray analysis, such as the Illumina HumanMethylation450 system, or by using methylation immunoprecipitation (e.g. using an anti-methylcytosine antibody) or treatment with a methylation-binding protein followed by microarray analysis or DNA sequencing, or by using methylation-sensitive restriction enzyme treatment followed by microarray or DNA sequencing, or by using methylation aware sequencing e.g. using a single molecule sequencing method (e.g. by a nanopore sequencing (Schreiber et al.

Proc Natl Acad Sci 2013; 110: 18910-18915) or by the Pacific Biosciences single molecule real time analysis (Flusberg et al. Nat Methods 2010; 7: 461-465)). Tissue-specific methylation levels can be measured in a same way. In yet other embodiments, other methods, for example, but not limited to targeted bisulfite sequencing, methylation-specific PCR, non-bisulfite based methylation-aware sequencing (e.g. by single molecule sequencing platforms (Powers et al. Efficient and accurate whole genome assembly and methylome profiling of E. coli. BMC Genomics. 2013;14:675)) can be used for the analysis of the methylation level of the plasma DNA for plasma DNA methylation deconvolution analysis.

[0084] At block 160, M values of a composition vector are determined. Each M value corresponds to a fractional contribution of a particular tissue type of the M tissue types to the DNA mixture. The M values of the composition vector can be solved to provide the N mixture methylation levels (e.g., methylation vector b) given the matrix A composed of NxM tissue-specific methylation levels (i.e. N tissue-specific methylation levels for each of the M tissue types). The M fractional contributions can correspond to a vector x that is determined by solving Ax=b. When N is greater than M, the solution can involve a minimization of errors, e.g., using least-squares.

[0085] At block 170, the composition vector is used determine an amount of each of the M tissue types in the mixture. The M values of the composition vector may be taken directly as the fractional contributions of the M tissue types. In some implementations, the M values can be converted to percentages. Error terms can be used to shift the M values to higher or lower values.

D. Applications

[0086] As mentioned above, the fractional contributions can be used in further measurements of the biological sample and other determinations, e.g., whether a particular chromosomal region has a sequence imbalance, whether a particular tissue type is diseased, and to determine which haplotype of two parental haplotypes is inherited by the fetus of a pregnant female from which the sample was obtained.

[0087] FIG. 2 shows a schematic diagram showing several potential applications of DNA methylation deconvolution (e.g., using plasma) according to embodiments of the present invention. In FIG. 2, a biological sample 205 is subjected to genome-wide bisulfite sequencing at 210. At

230, plasma DNA tissue mapping uses tissue-specific methylation profiles 220 to determine tissue contribution percentages. Example tissue-specific methylation profiles are shown as liver, blood cells, adipose tissues, lungs, small intestines, and colon. The contribution percentages can be determined as described above and elsewhere, e.g., solving Ax=b. Examples of applications include prenatal testing 241, cancer detection and monitoring 242, organ transplant monitoring, and organ damage assessment 244.

[0088] A list of methylation markers (genomic sites) that are useful for determining the contributions of different organs to the plasma DNA can be identified by comparing the methylation profiles (FIG. 2) of different tissues, including the liver, lungs, esophagus, heart, pancreas, sigmoid colon, small intestines, adipose tissues, adrenal glands, colon, T cells, B cells, neutrophils, brain and placenta. In various examples, whole genome bisulfite sequencing data for the liver, lungs, esophagus, heart, pancreas, colon, small intestines, adipose tissues, adrenal glands, brain and T cells were retrieved from the Human Epigenome Atlas from the Baylor College of Medicine (www.genboree.org/epigenomeatlas/index.rhtml). The bisulfite sequencing data for B cells and neutrophils were from the publication by Hodges et al. (Hodges et al; Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell 2011; 44: 17-28). The bisulfite sequencing data for the placenta were from Lun et al (Lun et al. Clin Chem 2013; 59:1583-94). In other embodiments, markers can be identified from datasets generated using microarray analyses, e.g. using the Illumina Infinium HumanMethylation450 BeadChip Array.

II. SELECTION OF METHYLATION MARKERS

[0089] Above, we have described the principle of using methylation analysis to determine the composition of a DNA mixture. In particular, the percentage contribution of different organs (or tissues) to the plasma DNA can be determined using methylation analysis. In this section, we further describe the method for the selection of methylation markers and clinical applications of this technology.

[0090] The results of determining the composition of the DNA mixture by methylation analysis are affected by the methylation markers used for the deconvolution of the composition of the DNA

mixture. Thus, the selection of appropriate genomic methylation markers can be important for the accurate determination of the constitution of the DNA mixture.

A. Criteria for a Methylation Marker for Deconvolution

[0091] For marker selection, the following three attributes may be considered. (i) It is desirable for a methylation marker to have a low variability in the methylation level measured in the same tissue type across different individuals. As the determination of the composition of the DNA mixture is dependent on the recognition of the tissue-specific methylation patterns, the low variability in methylation level in the same tissue type across different individuals would be useful for accurate identification of the tissue-specific patterns in the DNA mixture. In embodiments where the tissue-specific methylation levels are obtained from samples of other organisms (e.g., from a database), the low variability means that the methylation levels from the other samples are similar to the tissue-specific methylation levels for the current organism being tested.

[0092] (ii) It is desirable for a methylation marker to have a high variability in methylation levels across different tissues. For a particular marker, a higher difference in the methylation levels across different tissues can provide a more precise determination of the contribution of different tissues to the DNA mixture. In particular, an improvement in precision can be obtained by using one set of markers having attribute (ii) and another set of markers having attribute (iii).

[0093] (iii) It is desirable for a methylation marker to have a particularly different methylation level in a particular tissue when compared with those from most or all of the other tissues. In contrast to point (ii) above, a marker can have low variability in the methylation level of most tissues but its methylation level in one particular tissue is different from most of the other tissues. This marker would be particularly useful for the determination of the contribution of the tissue that has a different methylation level from other tissues.

B. Example

[0094] A principle of marker selection is illustrated in the following hypothetical examples in table 1.

	Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6
Liver 1	20%	69%	9%	9%	10%	90%

Liver 2	50%	70%	10%	10%	10%	90%
Liver 3	90%	71%	11%	11%	10%	90%
Heart	20%	20%	30%	13%	12%	12%
Lung	30%	30%	60%	17%	14%	84%
Colon	40%	40%	90%	20%	80%	80%

Table 1. Methylation densities in different tissues for 6 hypothetical methylation markers.

[0095] In this hypothetical example, marker 2 has lower variability in methylation density in the liver from three individuals when compared with marker 1. Therefore, marker 2 is superior to marker 1 as a signature for determining the contribution of the liver in a DNA mixture.

[0096] Compared with marker 4, marker 3 has a higher variability in methylation density across different tissue types. The same level of change in the estimated contribution from the different tissues would provide a bigger change in the deduced methylation density of the DNA mixture for marker 3 than for marker 4 according to the mathematical relationship discussed above. Therefore, the estimation of the contribution of each tissue can be more precise with marker 3.

[0097] Marker 5 has a low variability in methylation density across the liver, heart and lung. Their methylation densities vary from 10% to 14%. However, the methylation density of colon is 80%. This marker would be particularly useful for determining the contribution of the colon in the DNA mixture. Similarly, the heart is hypomethylated compared with the other tissues for marker 6. Therefore, the contribution of the heart can be accurately determined by marker 6. Thus, the combination of markers 5 and 6 would be able to accurately determine the contributions of the colon and the heart. The addition of markers 2 and 3 would then be sufficient to deduce the contribution of each of the four organs, including the liver, heart, lung and colon.

C. Different Types of Markers

[0098] A methylation marker may not necessarily need to have all of the above three attributes. A type I methylation marker would typically have attribute (iii) above. A number of such markers may also have attribute (i). On the other hand, a type II methylation marker would typically have attribute (ii) above. A number of such markers may also have attribute (i). It is also possible that a particular marker may have all three attributes.

[0099] In some embodiments, markers are broadly divided into two types (type I and type II). Type I markers have tissue specificity. The methylation level of these markers for a particular group of one or more tissues is different from most of the other tissues. For example, a particular tissue can have a significant methylation level compared with the methylation level of all the other tissues. In another example, two tissues (e.g., tissue A and tissue B) have similar methylation levels, but the methylation levels of tissues A and B are significantly different from those of the remaining tissues.

[0100] Type II markers have a high inter-tissue methylation variability. The methylation levels of these markers are highly variable across different tissues. A single marker in this category may not be sufficient to determine the contribution of a particular tissue to the DNA mixture. However, a combination of type II markers, or in combination with one or more type I markers can be used collectively to deduce the contribution of individual tissues. Under the above definition, a particular marker can be a type I marker only, a type II marker only, or be simultaneously both a type I and type II marker.

1. Type I Markers

[0101] In one embodiment, a type I marker can be identified by comparing the methylation density of the marker with the mean and standard deviation (SD) of methylation densities of this particular marker for all candidate tissues. In one implementation, a marker is identified if its methylation density in one tissue is different from the mean of all the tissues by 3 standard deviations (SD).

[0102] The methylation profiles of 14 tissues obtained from sources mentioned above were studied to select markers. In one analysis, a total of 1,013 type I markers were identified (markers labeled type I in Table S1 of Appendix A of U.S. Provisional Application No. 62/158,466) using the above criteria. In other embodiments, other cutoffs between the particular tissues and the mean methylation densities can be used, for example, but not limited to 1.5 SD, 2 SD, 2.5 SD, 3.5 SD and 4 SD. In yet another embodiment, a type I marker can identified through the comparison of the methylation density of the particular tissue to the median methylation density of all tissues.

[0103] In other embodiments, the type I markers can be obtained when more than one tissue (for example, but not limited to two, three, four or five tissues) show significantly different

methylation densities than the mean methylation density of all the candidate tissues. In one implementation, a cutoff methylation density can be calculated from the mean and SD of the methylation densities of all the candidate tissues. For illustration purpose, the cutoff can be defined as 3 SD higher or lower than the mean methylation densities. A marker is selected when the methylation densities of more than one (for example, but not limited to two, three, four, five, or more than five) tissues are more than 3 SD higher than the mean methylation density or more than 3 SD lower than the mean methylation density of the tissues.

2. Type II Markers

[0104] For identification of type II markers, the mean and SD of methylation densities across all 14 candidate tissues were calculated and the ratio of SD to the mean was denoted as the coefficient of variation (CV). In this illustrative example, we used a cutoff of >0.25 for the CV to identify the qualified type II markers, as well as the difference between the maximum and minimum methylation densities for the group of tissues exceeding 0.2. Using these criteria, 5820 type II markers were identified (markers labeled type II in Table S1 of Appendix A). In other embodiments, other cutoffs for the CV, for example but not limited to 0.15, 0.2, 0.3 and 0.4, can be used. In yet other embodiments, other cutoffs for the difference between the maximum and minimum methylation densities, for example, but not limited to 0.1, 0.15, 0.25, 0.3, 0.35, 0.4, 0.45 and 0.5, can be used.

[0105] In other embodiments, the average values across multiple samples of the same tissue type can be used to measure a variation of the methylation levels across different tissues. For example, 10 methylation levels of a same genomic site from 10 samples can be averaged to obtain a single methylation level for the genomic site. A similar process can be performed to determine average methylation levels for other tissue types for the genomic site. The average values across tissue types can then be used for determining whether the genomic site has significant variation across tissue types. Other statistical values can be used besides an average, e.g., a median or a geometric mean. Such statistical values can be used to identify type I and/or type II markers.

[0106] The different samples of a same tissue type (e.g., from different individuals) can be used to determine a variation of methylation levels across the different samples. Thus, if there are multiple samples of the same tissue type, embodiments can further measure the variation of a

particular marker amongst such samples of the same tissue type. A marker with a low variation across samples would be a more reliable marker than one with a high variation. Further details of markers and deconvolution can be found in commonly-owned U.S. Patent Publication 2016/0017419, entitled "Methylation Pattern Analysis Of Tissues In A DNA Mixture," by Chiu et al., and PCT Publication WO2014/043763 entitled "Non-Invasive Determination Of Methylome Of Fetus Or Tumor From Plasma."

D. Different Categories of Markers

[0107] A "category" for a genomic locus (methylation marker) corresponds to specific variation in methylation levels for a locus across different individuals for a same tissue type. Different categories can have different ranges of variation among a particular tissue type across individuals. A first category of methylation markers might have a difference of 10% in the methylation levels or lower among the individuals tested. A second category of methylation markers might have a difference of more than 10% in the methylation levels among the individuals tested. The use of methylation markers with low inter-individual variations (first category markers) would potentially improve the accuracy of determining the contribution of the particular organ in the DNA mixture.

E. Identification of Potential Methylation Markers

[0108] In some embodiments, potential methylation markers were identified in the following manner. Such potential methylation markers can then be subjected to the above criteria to identify type I and type II markers. In other embodiments, an identification of type I or type II is not needed. And, other embodiments may use other techniques to identify potential methylation markers.

[0109] In some embodiments, all CpG islands (CGIs) and CpG shores on autosomes were considered for potential methylation markers. CGIs and CpG shores on sex chromosomes were not used so as to minimize variation in methylation levels related to the sex-associated chromosome dosage difference in the source data. CGIs were downloaded from the University of California, Santa Cruz (UCSC) database (genome.ucsc.edu/, 27,048 CpG islands for the human genome) (Kent et al., The human genome browser at UCSC, Genome Res. 2002;12(6):996-1006) and CpG shores were defined as 2 kb flanking windows of the CpG islands (Irizarry et al. The

human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 2009; 41(2):178 - 186). Then, the CpG islands and shores were subdivided into non-overlapping 500 bp units and each unit was considered as a potential methylation marker.

[0110] The methylation densities (i.e., the percentage of CpGs being methylated within a 500 bp unit) of all the potential loci were compared between the 14 tissue types. As previously reported (Lun et al. Clin Chem. 2013; 59: 1583-94), the placenta was found to be globally hypomethylated when compared with the remaining tissues. Thus, the methylation profile of the placenta was not included at the marker identification phase. Using the methylation profiles of the remaining 13 tissue types, the two types of methylation markers were identified. For example, type I markers can refer to any genomic sites with methylation densities that are 3 SD below or above in one tissue when compared with the mean level of the 13 tissue types. Type II markers can considered highly variable when (A) the methylation density of the most hypermethylated tissue is at least 20% higher than that of the most hypomethylated one; and (B) the SD of the methylation densities across the 13 tissue types when divided by the mean methylation density (i.e. the coefficient of variation) of the group is at least 0.25. Lastly, in order to reduce the number of potentially redundant markers, only one marker may be selected in one contiguous block of two CpG shores flanking one CpG island.

F. Selection based on Application

[0111] The set of methylation markers chosen for particular applications can be varied depending on the parameters of the desired applications. For example, for applications focusing on haplotype or allele analysis, useful markers would be those located on the same cell-free DNA molecules as one of the heterozygous alleles. As cell-free DNA molecules (e.g., plasma DNA) are usually less than 200 bp, useful markers can be CpG sites within 200 bp of a heterozygous locus (e.g., a SNP). As another example, for applications in which the release of DNA from a particular tissue into plasma is of special significance, one can select a preferentially larger number of methylation markers that are differentially methylated in this tissue type (e.g. type I marker) when compared with the others in the marker set.

[0112] The number and choice of methylation markers in the deconvolution analysis can be varied according to the intended use. If the fractional contribution of the liver is of particular interest, e.g. in a patient who has received a liver transplant, more type I liver specific markers can be used in the deconvolution analysis to increase the precision of the quantification of the contribution of the transplanted liver to the plasma DNA.

III. COMPOSITION ACCURACY

[0113] As described above, embodiments can identify the tissue contributors of plasma DNA. In various examples, genome-wide bisulfite sequencing of plasma DNA was performed and analyzed with reference to methylation profiles of different tissues. Using quadratic programming as an example, the plasma DNA sequencing data were deconvoluted into proportional contributions from different tissues. Embodiments were tested for pregnant women, patients with hepatocellular, lung and colorectal carcinoma, and subjects following bone marrow and liver transplantation.

[0114] In most subjects, white blood cells were the predominant contributors to the circulating DNA pool. The placental contributions in pregnant women correlated with the proportional contributions as revealed by fetal-specific genetic markers. The graft-derived contributions to the plasma in the transplant recipients correlated with those determined using donor-specific genetic markers. Patients with hepatocellular, lung or colorectal cancer showed elevated plasma DNA contributions from the organ with the tumor. The liver contributions in hepatocellular carcinoma patients also correlated with measurements made using tumor-associated copy number aberrations.

[0115] In cancer patients and in pregnant women exhibiting copy number aberrations in plasma, methylation deconvolution pinpointed the tissue type responsible for the aberrations. In a pregnant woman diagnosed as having follicular lymphoma during pregnancy, methylation deconvolution indicated a grossly elevated contribution from B-cells into the plasma DNA pool and localized B-cells (instead of the placenta) as the origin of the copy number aberrations observed in plasma. Accordingly, embodiments may serve as a powerful tool for assessing a wide range of physiological and pathological conditions based on the identification of perturbed proportional contributions of different tissues into plasma.

A. Contribution of Different Types of Blood Cells

[0116] As an example of the methylation deconvolution, we determined the contribution of different tissues and cell types to the circulating DNA. Two blood samples were collected from two patients suffering from systemic lupus erythematosus (SLE). After collection, the venous blood samples were centrifuged at 1,500 g for 10 minutes. After centrifugation, the blood cells and the plasma were separated. DNA was then extracted from the blood cells. The DNA was bisulfite converted and sequenced using one lane of a flow cell in a HiSeq2000 sequencer. Two blood cell samples were analyzed using the cell-type-specific methylation pattern analysis. The methylation patterns of neutrophils, lymphocytes, the esophagus, colon, pancreas, liver, lung, heart, adrenal glands and hippocampus were included as potential candidates of the blood cell DNA. 609 methylation markers were selected for the analysis. The whole blood samples of the two subjects were also sent for cell counting to determine the fractional composition of the neutrophils and lymphocytes of the blood cells.

	Blood sample 1		Blood sample 2	
	Cell type-specific methylation pattern analysis	Blood cell counting	Cell type-specific methylation pattern analysis	Blood cell counting
Neutrophils	90.5%	93.6%	89.4%	89.9%
Lymphocytes	9.5%	6.4%	10.6%	10.1%
Esophagus	0%	-	0%	-
Colon	0%	-	2%	-
Pancreas	0%	-	0%	-
Liver	0%	-	1%	-
Lung	1%	-	1%	-
Heart	0%	-	3%	-
Adrenal gland	0%	-	0%	-
Hippocampus	0%	-	0%	-

Table 2. Blood tissue contributions by deconvolution pattern analysis and cell counting

[0117] For methylation pattern analysis, neutrophils and lymphocytes were determined as the major components constituting the blood cell DNA. The relative proportion of the contribution of neutrophils and lymphocytes resemble their relative abundance in the blood samples according to the cell counting analysis.

B. Pregnant Women

[0118] The contributions of different tissues, including the liver, lung, pancreas, colon, hippocampus, small intestines, blood cells, heart, adrenal gland, esophagus and placenta, were analyzed using methylation analysis of the plasma DNA of pregnant women. As the placental genotype is in general identical to the fetus's genotype but different from the pregnant woman's genotype, the precise contribution of the placenta to the maternal plasma can be accurately determined by counting the number of fetal specific-alleles in the sample.

1. Composition and Correlation to Fetal DNA Percentage

[0119] Genome-wide bisulfite sequencing of plasma DNA was performed for 15 pregnant women, five from each of first, second and third trimesters. Methylation deconvolution was performed and the percentage contributions from different tissues were deduced. The contributions of different organs were determined based on the methylation levels (such as methylation densities) of all the type I and type II markers in table S1 using quadratic programming analysis.

[0120] FIG. 3A shows a graph 300 of percentage contributions of different organs to plasma DNA for 15 pregnant women according to embodiments of the present invention. Each bar corresponds to the results of one sample. The different colors represent the contributions of different organs into plasma. These results show that the white blood cells (i.e. neutrophils and lymphocytes) are the most important contributors to the plasma DNA pool. This observation is consistent with those previously obtained following bone marrow transplantation (Lui YY et al. Clin Chem 2002; 48: 421-7).

[0121] FIG. 4 shows a table 400 of percentage contributions determined from a plasma DNA tissue mapping analysis among pregnant women according to embodiments of the present invention. These results also show that the placenta is another key contributor of the plasma DNA in pregnant women, with fractional concentrations from 9.9% to 38.4%.

[0122] We also measured the placental contributions using paternally-inherited fetal single nucleotide polymorphism (SNP) alleles that were not possessed by the pregnant women as previously described (31). To analyze the fetal-specific SNP alleles, the genotypes of the fetuses were determined by analyzing the chorionic villus samples or the placenta. The genotypes of the

pregnant women were determined by analyzing the blood cells. The SNP-based results show the independent validation of the methylation deconvolution results.

[0123] FIG. 3B shows a plot 350 of a correlation between the plasma DNA fractions contributed by the placenta deduced from plasma DNA methylation deconvolution and the fetal DNA fractions deduced using fetal-specific SNP alleles according to embodiments of the present invention. Plot 350 shows that the placental contributions determined by methylation deconvolution have a strong correlation with the fetal DNA fractions measured using SNPs ($r = 0.99$, $p < 0.001$, Pearson correlation). Accordingly, a good positive correlation is observed between the values of the two parameters, suggesting that the plasma DNA methylation deconvolution accurately determines the contribution of the placenta to the maternal plasma samples.

[0124] FIG. 5 shows plots of percentage contributions of organs other than the placenta by plasma DNA tissue mapping and fetal DNA fractions based on fetal-specific SNP alleles according to embodiments of the present invention. The X-axis represents the fetal DNA fractions estimated by SNP-based analysis and the Y-axis represents the percentage contribution deduced by plasma tissue DNA mapping analysis. Plasma DNA contributions of the neutrophils show a reverse correlation. This is likely due to the fact that neutrophils are a major contributor to the plasma DNA pool and hence, as the placental contribution increases, the relative contribution from the neutrophils would by necessity decrease. The methylation deconvolution results of the remaining tissues show no correlation with the fetal DNA fraction.

[0125] FIG. 6 shows a table 600 of percentage contributions from plasma DNA tissue mapping analysis among the non-pregnant healthy control subjects according to embodiments of the present invention. When the process was applied to plasma of non-pregnant healthy controls, placental contribution was absent in most samples (median: 0%; interquartile range: 0% to 0.3%).

2. Comparison of Selected Markers vs. Random Markers

[0126] The accuracy of the percentage contributions was tested with select markers relative to random markers. Different composition calculations were done for different sets of markers. One set was chosen based on criteria mention above, and the other was a random set. The results show

that it is important to judicially choose the methylation markers (genomic loci) use, in order to obtain accurate results.

[0127] Eleven pregnant women and four healthy non-pregnant subjects were recruited for this analysis. Their plasma DNA was bisulfite converted and sequenced using the Illumina HiSeq2000 sequencer. Each plasma sample was sequenced with one lane of a sequencing flow cell. The sequence reads were then analyzed using a bioinformatic program, Methy-Pipe (Jiang P. PLoS One 2014; 9: e100360). This program can align the bisulfite-converted sequence reads to the reference genome and determine the methylation status of each CpG site on each sequenced fragment.

[0128] The first set of markers have high specificity for identifying the different tissues in the plasma DNA. For each tissue type, markers that have the biggest difference in methylation density compared with the other tissues were selected. The markers were determined from genomic regions containing at least one CpG dinucleotide. In this example, CpG islands (CGIs) were used as potential markers, having a high frequency of CpG sites in a particular stretch of DNA. CGIs in this particular example are downloaded from the University of California, Santa Cruz (UCSC) database: (genome.ucsc.edu). In total, we obtained 27,048 CpG islands from the human genome. The median size of a CpG island is 565 bp (range: 200 bp to 45 kb). 90% of the islands are less than 1.5 kb.

[0129] For each methylation marker, the difference in methylation density between the tissue-of-interest and the other tissues was determined. The difference is then expressed as the number of standard deviations (SDs) across the other tissues. For the tissue-of-interest, all the markers were ranked according to this difference in methylation density. The 20 markers with the biggest difference above (10 markers) and below (10 markers) the mean methylation densities of the other tissues were selected. The number of markers can vary, for example, but not limited to 5, 15, 20, 30, 40, 50, 100 and 200.

[0130] In addition, markers with a high variability across all the different tissues were also selected. In this example, markers with >50% difference between the tissues with the highest and lowest methylation densities were selected. In other applications, other values, for example, but not limited to 20%, 30%, 40%, 60%, 70% and 80%, can be used. Furthermore, the variability of methylation densities across different tissues was also calculated based on the mean and SD. In

this example, a marker was also selected if the value of SD is more than two times the mean. In other applications, other cutoff values, for example, but not limited to 1, 1.5, 2.5 and 3, can also be used. Based on these selection criteria, 344 methylation markers were selected for the first set.

[0131] For the second set, 341 markers were randomly selected from the 27,048 CGIs discussed above. All the CGIs were first numbered from 1 to 27,048. Then a random number (between 1 and 27,048) was generated by a computer for marker selection. This process was then repeated until a total of 341 markers were selected. If a random number generated had been used, another one would be generated. This set of markers is expected to have a much lower specificity in identifying the tissue-specific methylation patterns. Thus, the accuracy of determining the composition of the plasma DNA is expected to be reduced.

[0132] FIG. 7 shows a table 700 of the estimated contributions of different organs to the plasma DNA for 11 pregnant women and 4 non-pregnant healthy subjects using the first set of markers (with high organ specificity) according to embodiments of the present invention. The fetal DNA fractions were determined by counting fetal-specific alleles and are shown in the bottom row. In each of the four non-pregnant control subjects, the contribution of the placenta to the plasma was determined to be close to 0%. This indicates the specificity of this approach.

[0133] FIG. 8 shows a table 800 of the estimated contributions of different organs to the plasma DNA for 11 pregnant women and 4 non-pregnant healthy subjects using the second set of markers (with low organ specificity) according to embodiments of the present invention. The fetal DNA fractions determined by counting fetal-specific alleles are shown in the bottom row. Using these less specific markers, a relatively non-concordant percentage of contribution from the placenta was observed, and considerable contributions from the placenta were observed in the four non-pregnant control subjects. This indicates that the tissue specificity of the markers is important in this approach.

[0134] FIG. 9A is a plot 900 showing the correlation between the estimated fetal DNA fraction (contribution from the placenta) and the fetal DNA fraction determined by counting the fetal-specific alleles in the maternal plasma samples. The results from the two techniques have good correlation using the first set of methylation markers. However, using the second set of methylation markers, the estimation by using the methylation analysis showed significant deviation from the true values determined using fetal-specific alleles counting.

[0135] FIG. 9B is a plot 950 showing absolute difference between the estimation from methylation markers and fetal DNA fraction determined by fetal-specific alleles counting. The median error of the estimation using methylation analysis were 4% and 8% using the first set of markers and the second set of markers, respectively.

C. Effect of Different Criteria

[0136] As described above, various criteria can be used to identify markers of different types. For example, a type I marker can be identified by a methylation level in a particular tissue that is different from the mean methylation level for all tissues, e.g., at least by a specific threshold, such as 3 SD. And, for type II markers, criteria of a certain variation and maximum difference are used. Sections below show accuracy of different criteria for identifying markers.

1. Performance of Markers with Less Stringent Criteria

[0137] We compared the performance of methylation deconvolution analysis using markers with different variability across different tissues. The placental contributions to plasma DNA were determined for 15 pregnant women based on two sets of markers with different selection criteria. Both sets of markers include all the type I markers as described in previous sections. However, the selection criteria of type II markers are different for the two sets of markers.

[0138] Set I markers include all the 5,820 type II markers fulfilling the criteria of having methylation density CV >0.25 and the difference between the maximum and minimum methylation densities for the groups of tissues exceeding 0.2. For Set II markers, the CV requirement was >0.15 and the difference between the maximum and minimum methylation densities for the groups of tissues exceeded 0.1. There were 8,511 type II markers in this set of markers.

[0139] FIG. 10A is a graph 1000 showing placental contribution to plasma DNA deduced using markers with different selection criteria according to embodiments of the present invention. The vertical axis corresponds to placental contribution deduced using the set II markers. The horizontal axis corresponds to placental contribution deduced using the set I markers. There was a good correlation between the placental contribution results based on the two sets of markers with different selection criteria (r=0.99, Pearson correlation). Accordingly, good accuracy can be

obtained using the requirements of CV >0.15 and of the difference between the maximum and minimum methylation densities for the groups of tissues exceeding 0.1.

2. Effect of Methylation Level Variation within Same Type of Tissue

[0140] To investigate if the variation in methylation level of markers between the same type of tissues (e.g. from different individuals) would affect the performance of deconvolution analysis, we analyzed placental tissues from two pregnant cases. Two categories of methylation markers were identified. Specifically, the two categories were identified based on their similarity in methylation levels in two placental tissues. Markers of category i have a methylation density of 10% or lower. Markers of category ii have high variability between the two placental tissues (difference in methylation density of more than 10%).

[0141] FIG. 10B is a graph 1050 showing the accuracy of plasma DNA deconvolution using markers with low variability (category i) and high variability (category ii) in the same type of tissue. Plasma DNA deconvolution was performed to determine the placental contribution to the plasma DNA for 15 pregnant women. For each marker, the mean of the methylation densities of the two placental tissues were used to represent the methylation level of the placenta in the analysis. For each of the deconvolution analysis using the category i and category ii markers, a total of 1,024 markers were used.

[0142] The amount of placentally-derived DNA in plasma was further determined based on the proportion of the fetal-specific SNP alleles. The percentage contribution deduced by the methylation deconvolution analysis based on category i and category ii markers were then compared with the results based on fetal-specific SNP alleles. The median deviation of the derived placental contribution from the value estimated based on fetal-specific alleles was 2.7% and 7.1% using category i and category ii markers, respectively. Thus, the use of category i markers which had lower inter-individual variation in the tissue methylation level gave better accuracy in the methylation deconvolution analysis.

[0143] Significantly higher difference between the results from methylation deconvolution and fetal-specific allele analysis was observed when markers with high variability within the same type of tissue (category ii) were used ($P<0.0001$, Wilcoxon sign-rank test). In other words, the use of markers with low variability within the same type of tissue would increase the accuracy of

methylation deconvolution analysis. Accordingly, markers can be selected based on the variability within the same type of tissues, for example, but not limited to the value of CV and the difference between the maximum and minimum methylation density for the same type of tissues.

IV. DECONVOLUTION OF FETAL SIGNATURES

[0144] If a genomic signature (e.g., a particular SNP allele) is known, embodiments can determine which tissue is the origin of such signatures. Thus, if a particular signature is representative of a fetus (e.g., a paternal allele at a particular locus), then the fractional contribution for the signature would be substantial for the placental tissue.

[0145] To illustrate that single nucleotide alteration can also be used to determine the tissue of origin that the alteration is derived from, we analyzed the plasma DNA of a pregnant woman. The placenta and the maternal buffy coat were genotyped to identify the SNPs that the mother was homozygous and the fetus was heterozygous. We denote the allele shared by the fetus and the mother as A and the fetal-specific allele as B. Therefore, the mother had a genotype of AA and the fetus had a genotype of AB at each of these SNPs.

[0146] After bisulfite sequencing of the maternal plasma DNA, all DNA fragments carrying the fetal-specific allele (B allele) and at least one CpG site were selected and used for downstream analysis. A total of 1.31 billion fragments were sequenced and 677,140 fragments carrying the fetal-specific allele (B allele) were used for the deconvolution analysis. All CpG sites that were covered by at least 10 DNA fragments were used for deconvolution analysis. Other numbers of DNA fragments covering a site can be used, such as 5, 15, 20, 25, or 30. As the B allele was fetal specific, these DNA fragments were expected to be derived from the placenta.

Tissue	Contribution (%)
Liver	0.9
Lung	0.0
Colon	0.0
Small intestines	0.0
Pancreas	0.5
Adrenal glands	0.0
Esophagus	3.1
Adipose tissues	0.0
Heart	0.0
Brain	0.3

T cells	0.0
B cells	0.0
Neutrophil	0.0
Placenta	95.2

Table 3. Methylation deconvolution analysis using fetal-specific allele.

[0147] In Table 3, from the methylation deconvolution analysis, it was shown that the placenta was deduced to be the major contributor for these DNA fragments carrying fetal-specific SNP alleles. These results suggest that the methylation deconvolution analysis accurately identified the tissue origin of these DNA fragments carrying fetal-specific alleles.

[0148] This shows that a particular allele can be attributed to a fetus. Such a technique is described in more detail below for determining genotypes and haplotypes of a fetus using methylation deconvolution analysis.

V. DETERMINATION OF FETAL GENOME (MUTATIONAL ANALYSIS)

[0149] For noninvasive prenatal testing, the analysis of the inheritance of a maternal mutation using maternal plasma DNA is a challenging task. For example, if a pregnant woman is heterozygous for a mutation, the analysis on the mutational status of the fetus using maternal plasma DNA analysis would be technically difficult because both the mutant and the normal alleles would be present in her plasma, regardless of the mutational status of her fetus. Previously, a number of different approaches have been developed to address this problem (Lun et al. Proc Natl Acad Sci U S A. 2008;105:19920-5; Lo et al. Sci Transl Med. 2010;2:61ra91; Lam et al. Clin Chem. 2012;58:1467-75). The principle of these previous approaches involves the comparison between the relative amounts of the mutation and the normal allele in maternal plasma. To enhance the statistical power of the comparisons, some of these approaches further involve the comparisons of the relative amounts of SNP alleles linked to the mutation and those linked to the normal allele. As an alternative or in addition, some embodiments of the present invention can deduce the mutational status of the fetus by methylation deconvolution analysis.

A. Contribution for Alleles using Methylation Deconvolution

[0150] In this example, a genotype of the fetus is determined. Assume the genotypes of the father and mother are NN and MN at a particular locus, respectively. M and N denote the mutant

and normal alleles, respectively. In this scenario, the fetus can inherit either the M allele or the N allele from the mother. Therefore, there are two possible genotypes for the fetus, namely MN and NN. In maternal plasma, DNA carrying the fetal genotype is actually derived from the placenta. Thus, these DNA fragments would exhibit the placental methylation profile.

[0151] FIG. 11A shows a first scenario where the fetus has inherited the M allele from the mother and has a genotype of MN at a particular locus according to embodiments of the present invention. In the top part of FIG. 11A (labeled Genotypes), the father is shown as having genotype NN, the mother is shown as having genotype MN, and the fetus is shown as having genotype MN. The DNA fragments that exhibit the placental methylation profile are marked with a P, where are shown on the fetal genotype. For example, the placental methylation profile can correspond to certain methylation levels at genomic sites near the particular locus. DNA fragments that align to the particular locus can also include genomic sites near the locus (e.g., within 200 bp of the locus), and thus can be used to measure methylation levels for the methylation deconvolution analysis. Considering the genotypes of the parents, the M allele is specific for the mother and the N allele is shared between the father and the mother.

[0152] In the bottom part of FIG. 11A (labeled Maternal plasma), instances of the two alleles M and N are shown, with each instance representing a different DNA molecule in the plasma at the locus of interest. Only a small number of DNA molecules are shown for illustration purposes. In this example, the fetal DNA percentage is assumed to be 25%, as shown by 25% of the DNA molecules being marked with a P.

[0153] In the maternal plasma sample, we selectively analyzed the DNA fragments carrying the M allele and performed the methylation deconvolution analysis. Because the fetus has a genotype of MN, the placenta would contribute both M and N alleles to the maternal plasma DNA. Therefore, some of the DNA fragments carrying the M allele would also carry the placenta-specific methylation profile at genomic sites near the locus. The methylation deconvolution analysis would indicate that some of the DNA fragments carrying the M allele would be derived from the placenta, and thus the fetal genotype does include the M allele.

[0154] FIG. 11B shows a second scenario where the fetus has inherited the N allele from the mother and has the genotype of NN at a particular locus according to embodiments of the present invention. In this situation, only the DNA fragments carrying the N allele would exhibit the

placental methylation profile in maternal plasma. Thus, the selective analysis of the DNA fragments carrying the M allele with methylation deconvolution would indicate that these DNA fragments do not have significant contribution from the placenta. Accordingly, it can be determined that the fetus does not have M, and thus has a genotype of NN.

[0155] In some embodiments, the placental contribution for the M and N alleles can be compared. Here, we assume that the fetal DNA accounts for approximately 10% of the total maternal plasma DNA. The selective deconvolution of the M and N alleles would be useful to indicate which allele the fetus has inherited from the mother. The expected results are shown in Table 4 below:

	Fetal genotype	
	MN	NN
Placental contribution for plasma DNA carrying the M allele	Approximately 10%	Insignificant (close to zero)
Placental contribution for plasma DNA carrying the N allele	Approximately 10%	Approximately 20%
Ratio of placental contribution for M and N alleles (M:N)	1:1	0:2

Table 4. Placental contributions for M and N alleles for NN paternal genotype.

[0156] In Table 4, the percentage placental contribution of the M and N alleles can be compared. An approximately equal placental contribution for the two alleles (e.g., within a threshold of each other) suggests that the fetal genotype is MN. On the other hand, a significantly higher placental contributions for the N allele compared with the M allele would indicate a fetal genotype of NN.

[0157] In another embodiment, the paternal genotype does not need to be taken into account. In this situation, the possible genotypes of the fetus include MM, MN, and NN.

	Fetal genotype		
	MN	NN	MM
Placental contribution for plasma DNA carrying the M allele	Approximately 10%	Insignificant (close to zero)	Approximately 20%
Placental contribution for plasma DNA carrying the N allele	Approximately 10%	Approximately 20%	Insignificant (close to zero)

Ratio of placental contribution for M and N alleles (M:N)	1:1	0:2	2:0

Table 5. Placental contributions for M and N alleles for unknown paternal genotype.

[0158] In table 5, the placental contribution for the DNA fragments carrying the M and N alleles for different fetal genotypes are shown. When the fetus has a genotype of MM, the placental contribution for the M allele would be significantly higher than that for the N allele. When the fetus has a genotype of NN, the placental contribution for the N allele would be significantly higher than that for the N allele. When the fetus has a genotype of NM, the placental contribution for the M allele would be approximately equal to the placental contribution for the N allele.

[0159] Accordingly, where the paternal genotype is not known, fractional contributions can be determined for both alleles. That is, a first fractional contribution can be determined using a first set of cell-free DNA molecules that align to the locus and include N. Methylation levels of the first set of cell-free DNA molecules can be measured at K genomic sites near the locus. And, a second fractional contribution can be determined using a second set of cell-free DNA molecules that align to the locus and include M. Methylation levels of the second set of cell-free DNA molecules can be measured at the K genomic sites near the locus. For the first scenario of the fetal genotype being MN, the fractional contributions determined for either allele would be about the same, as can be tested to by determining whether the fractional contributions are within a threshold value of each other.

[0160] To illustrate the feasibility of this approach, we analyzed the plasma DNA of a pregnant woman. The plasma DNA was bisulfite converted and analyzed using massively parallel sequencing. In addition, the placenta and blood cells were analyzed to determine the genotype of the fetus and the mother. For illustration purposes, a SNP located within the *KLF2* gene was analyzed. For this SNP, the genotypes of the mother and the fetus were CG and CC, respectively. With this genotype combination, the placenta would contribute the C allele to the maternal plasma, but all the G alleles in the maternal plasma would be derived from the maternal tissues.

[0161] In the sequencing data, there were 24 fragments carrying the G allele and 55 fragments carrying the C allele. The CpG sites within these DNA fragments were used for methylation deconvolution. In this analysis, an objective is to determine the placental contribution of the two

alleles. To illustrate the principle, only the placenta and the blood cells were considered as candidate tissues for the methylation deconvolution analysis. In another embodiment, three or more types of tissues can be used as candidates. In yet another embodiment, tissues expected to have significant contribution, for example blood cells, liver, lung, intestines and placenta, can be used as candidates.

	C allele	G allele	C/G ratio
Placenta	62.6%	1.8%	34
Blood cells	37.4%	98.2%	

Table 6. Placental contributions for C and G alleles for unknown paternal genotype.

[0162] In Table 6, the contribution from the placenta was deduced to be 62.6% and 1.8% for the C allele and the G allele, respectively. The ratio of placental contribution for C/G is 34. These results suggest that the genotype of the fetus would be CC. This is consistent with the genotyping result of the placental tissue.

[0163] This embodiment is different from and potentially has more utility than a previous method for noninvasive prenatal testing based on the analysis of allelic ratio for DNA with a specific methylation pattern (Tong et al. Clin Chem 2006; 52: 2194-202). In this previous method, tissue-specific DNA is first identified from a DNA mixture (e.g. plasma DNA) based on methylation pattern. For example, a particular gene is completely unmethylated in the blood cells and methylated in the placenta. The identification is performed using an enzyme that leaves the methylated placental DNA intact.

[0164] Thus, all the methylated DNA molecules remaining in the plasma would be derived from the placenta rather than from the blood cells. Then, the allelic ratio for a SNP located on the placenta-derived DNA molecules can be determined by measuring amounts of the different alleles at the locus using the intact placental DNA. When the fetus is heterozygous for the SNP, the ratio of the two alleles in the placenta-specific DNA would be approximately 1. However, if the fetus is affected by an aneuploid chromosome and has three copies of the chromosome carrying this particular SNP, the ratio of the two alleles would be either 1:2 or 2:1.

[0165] In this previous method, the tissue-specific DNA molecules need to be first identified based on a methylation status that is unique to the tissue of interest. The methylated DNA molecules are unique for the placenta because the blood cells are completely unmethylated for the

targeted region. However, in this present embodiment, the uniqueness of a certain methylation state is not required. The candidate tissues only need to be different in their methylation profiles, accordingly more loci can be used, thereby enabling haplotype deconvolution. Thus, the tissue contributions can be determined for the different alleles based on their methylation profiles. Further, the previous method may be more susceptible to statistical variations as the numbers of fetal reads with each allele are compared directly to each other. Whereas, when the placental contributions are compared to each other, the numbers of fetal reads are not compared directly to each other. Instead, the placental contribution is determined from all of the reads (methylated or not), and thus the placental contributions can be the same, even when the number of fetal reads differs. Thus, a coverage bias to one haplotype can be accounted for.

B. Determination of Inherited Haplotype using Deconvolution

[0166] It has previously been demonstrated that through the analysis of plasma DNA (or other cell-free DNA) of a pregnant woman carrying a fetus, the maternal haplotypes inherited by the fetus can be deduced using the process of relative haplotype dosage analysis (RHDO) (Lo et al. Sci Transl Med 2010; 2: 61ra91 and U.S. Patent 8,467,976). In this method, one uses the haplotype information for the pregnant woman. This latter information can be obtained using family analysis or a method for the direct analysis of the haplotype (e.g. Fan et al. Nat Biotechnol 2011; 29: 51-57; Snyder et al. Nat Rev Genet 2015; 16: 344-358). SNPs that are heterozygous in the mother but homozygous in the father can be used for the RHDO analysis. Such a use of specific SNPs can limit the loci that can be used, and therefore limit the amount of data and accuracy. Embodiments may not be so restricted to such specific SNPs. Further, embodiments can be used in combination with the above references to provide increased accuracy.

[0167] Embodiments can use methylation deconvolution to determine placental contributions using the cell-free DNA molecules for two haplotypes. The placental contributions can be compared to determine which haplotype is inherited by the fetus. Embodiments can start with deduced maternal or paternal haplotypes, and then measure the methylation levels of plasma DNA molecules containing SNP alleles in each of those deduced haplotypes. One can then perform methylation deconvolution. The fetal haplotype can be identified as the one with the highest placental contribution from the methylation deconvolution analysis. In all of the above embodiments, the paternal or maternal haplotypes can, instead of being a deduced one, also be

determined by family analysis (i.e. by analyzing the DNA of other family members) or by a direct method (e.g. the method described by Fan et al Nat Biotechnol 2012).

1. Maternal Haplotypes

[0168] In this example, we demonstrate plasma DNA methylation deconvolution analysis can be used for deducing the maternal haplotypes inherited by an unborn fetus. A source of genomic DNA from the pregnant woman, e.g. the buffy coat DNA, can be subjected to genotyping, e.g. using a microarray. Then, the maternal genotyping results are entered into a haplotype deduction program (e.g., IMPUTE2, Howie et al. PLoS Genet. 2009;7:e1000529) to deduce the likely first maternal haplotype and the second maternal haplotype. Population-specific genotype and haplotype information can be taken into consideration for improving the accuracy of deduction. In other embodiments, the parental haplotypes can be worked out by single molecule analysis, for example but not limited to the methods described by Fan et al (Nat Biotechnol. 2011;29:51-7) , Kaper et al (Proc Natl Acad Sci U S A. 2013;110:5552-7), Lan et al. (Nat Commun. 2016;7:11784) and Selvaraj et al (Nat Biotech 2013;31:1111–1118). Then, maternal plasma DNA can be subjected to genome-wide bisulfite sequencing and alignment to reference genomic sequences. Methylation deconvolution can then be performed for each of the predicted haplotypes. As fetal DNA in maternal plasma is predominantly of placental origin, the maternal haplotype inherited by the fetus is the one that shows the highest placental contribution.

[0169] The maternal haplotype information can be used to link the SNP alleles and the CpG sites on the same homologous chromosome together. Then, DNA fragments from the same chromosome copy (haplotype) can be identified using the SNP alleles. The CpG sites (or other sites) on this particular chromosome copy (haplotype) can be used for the methylation deconvolution. As the number of CpG sites that can be used for deconvolution would be proportional to the number of SNPs on the homologous chromosome and much bigger than the number of CpG sites linked to a single SNP in the haplotype-based deconvolution analysis, this method would be more precise than the deconvolution analysis using CpG site(s) that are linked to a single SNP. The principle is illustrated in FIG. 12A.

[0170] FIG. 12A shows a determination of a maternal haplotype inherited by a fetus using methylation deconvolution according to embodiments of the present invention. In the top part of

FIG. 12A, the two haplotypes of the mother and the fetus are shown at three loci that the mother is heterozygous. The two maternal haplotypes are labeled as Hap I and Hap II. In this example, the fetus has inherited Hap I from the mother. For illustration purpose, only the SNP loci that the mother is heterozygous are shown. For illustration purposes, the father is homozygous for each of these loci in this example. However, the same principle extends to scenarios that the father is heterozygous without any change.

[0171] In the bottom part of FIG. 12A (labeled Maternal plasma), instances of the two alleles at each locus are shown, with each instance representing a different DNA molecule in the plasma at the locus of interest. Only a small number of DNA molecules are shown for illustration purposes. In this example, the fetal DNA percentage is assumed to be 20%, as shown by 20% of the DNA molecules being marked with a P.

[0172] In maternal plasma, the DNA molecules carrying the fetal genotype are derived from the placenta and therefore carrying the placental specific methylation patterns. The circles labeled with "P" represent the CpG sites exhibiting the placental methylation pattern near the heterozygous locus. A read including a heterozygous locus and a neighboring site can be used for measuring a methylation level for detecting the placental methylation pattern. In this example, an objective is to determine if the fetus has inherited Hap I or Hap II from the mother. To achieve this, plasma DNA fragments that carry alleles on Hap I and cover at least one CpG site are selected for methylation deconvolution. As the fetus has inherited Hap I from the mother, the placenta would contribute a significant proportion to this pool of plasma DNA molecules. On the other hand, when the fragments carrying alleles on Hap II are analyzed with methylation deconvolution, a very low contribution from the placenta would be observed.

[0173] To illustrate this, we analyzed the maternal plasma sample stated above for Table 6. We focused on a 5-Mb region on chromosome 1. SNPs where the mother was heterozygous and the fetus was homozygous were selected for the analysis. For each of these SNP loci, the alleles that were shared between the mother and the fetus formed one haplotype (denoted as Hap I) and the alleles that were present only on the maternal genome formed another haplotype (denoted as Hap II). Thus, in this example, there are two maternal haplotypes (Hap I and Hap II) and the fetus has inherited Hap I from the mother. In the maternal plasma, the DNA fragments carrying the alleles on Hap I and those carrying the alleles on Hap II were analyzed separately using methylation

deconvolution. All the CpG sites on the same plasma DNA molecule of a heterozygous SNP were used for the deconvolution analysis. In this example, none of these CpG sites overlapped with type I or type II markers.

	Hap I	Hap II
Liver	0%	0%
Lung	0%	6.7%
Colon	3.4%	6.2%
Small intestines	0%	10.6%
Pancreas	4.1%	0%
Adrenal glands	0%	4.6%
Esophagus	0%	0%
Adipose tissues	3.7%	3.6%
Heart	0%	0%
Brain	6.8%	10.6%
T cells	6.8%	21%
B cells	8.9%	11.7%
Neutrophil	12.7%	25%
Placenta	53.5%	0%

Table 7. Methylation deconvolution for Hap I and Hap II.

[0174] Table 7 shows deconvolution of plasma DNA fragments carrying the alleles on the two maternal haplotypes, namely Hap I and Hap II. The fetus had inherited the maternal Hap I. From this deconvolution analysis, the placenta was deduced to contribute 53.5% of the plasma DNA fragments carrying the alleles on Hap I. On the other hand, there was no contribution from the placenta to the plasma DNA fragments carrying the alleles on Hap II. Therefore, the methylation deconvolution analysis had accurately predicted that the fetus had inherited Hap I from the mother. Greater accuracy may be achieved using CpG sites that overlap with type I and/or type II markers.

[0175] As a further example, to demonstrate the practical utility of this approach, another pregnant woman was recruited. Maternal peripheral blood was taken. The blood sample was fractionated into plasma and the cellular components. The maternal buffy coat was analyzed using an Illumina HumanOmni2.5-8 BeadChip array. We used IMPUTE2 (Howie et al. PLoS Genet. 2009;7:e1000529) to deduce the phase of 851 heterozygous SNPs on a 5 Mb region on the

telomeric end of chromosome 1p. The haplotype phasing was based on reference haplotypes of 1000 genomes (mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.tgz).

[0176] After the phased haplotypes were obtained, the CpG sites linked to the two haplotypes were used to perform methylation deconvolution. All the CpG sites on the same plasma DNA molecule of a heterozygous SNP were used for the deconvolution analysis. In this example, none of these CpG sites overlapped with type I or type II markers. Among the 851 SNPs used for the deconvolution, 820 (96.2%) were on intron and intergenic regions. None of them overlapped with CpG islands or shores.

	Hap I	Hap II
Liver	0	0
Lung	0	5.4
Colon	0	6.2
Small intestine	0	0
Pancreas	0	25
Adrenal glands	0	0
Esophagus	0	0
Adipose tissues	0	17.8
Heart	0	0
Brain	0	0
T cells	11	7.9
B cells	0	0
Neutrophils	20.2	28.4
Placenta	68.9	9.3

Table 8. Methylation deconvolution for Hap I and Hap II.

[0177] Table 8 shows a deconvolution of plasma DNA fragments carrying the alleles on the two maternal haplotypes deduced from a panel of reference haplotypes. The two haplotypes are named Hap I and Hap II. The deduced Hap I has significantly higher amount of placental contribution than Hap II, namely, 68.9% versus 9.3%. Thus, the maternal Hap I was deduced to have been inherited by the fetus. The maternal inheritance relied on haplotype deduction was consistent with results from maternal and fetal genotypes.

[0178] An advantage of this method is that one is not restricted to SNPs for which the father of the fetus is homozygous and the mother of the fetus is heterozygous. Indeed, in the above example,

we had performed the analysis without knowing or deducing the paternal genotype or haplotype. This is an advantage over previously described methods ((Lo et al. Sci Transl Med 2010; 2: 61ra91, US Patent 8,467,976, Fan et al. Nature 2012; 487: 320- 324, Kitzman et al. Sci Transl Med 2012; 4: 137ra76).

[0179] In some embodiments, a first fractional contribution for a first haplotype can be compared to a reference value derived based on the fetal DNA fraction to determine whether the haplotype has been inherited by the fetus. The cutoffs can be calculated as for example but not limited to 1 time, 1.2 times, 1.4 times, 1.6 times, 1.8 times, 2 times, 2.2 times, 2.4 times, 2.6 times or 2.8 times of the fetal DNA fraction. In this manner, the second fractional contribution for a second haplotype does not need to be determined, if the first fractional contribution is sufficiently large.

[0180] In some embodiments, the inherited haplotype may have a deconvoluted fractional concentration double that of the fetal fraction and the non-inherited one has an insignificant contribution. The contribution of the non-inherited haplotype may not have a zero contribution as the paternal haplotypes may give noise to this analysis because some paternal alleles may be the same as the maternal alleles. If the level of noise is high, the fractional contribution of the second haplotype can be determined, and the one with a higher deconvoluted fraction can be deduced to be inherited by the fetus.

[0181] Some implementations could test both haplotypes using the reference value, to confirm that only one is inherited. If both appear to be inherited, then the two fractional contributions can be compared to each other. Additionally, if both appear to be inherited, the paternal genome can be checked, as the fetus could have inherited a paternal haplotype that matches the non-inherited maternal haplotype.

[0182] In other embodiments, the second fractional contribution can be used to determine the reference value, e.g., the second fractional contribution plus a threshold value. Thus, the reference value can be a sum of the second fractional contribution and a threshold value.

2. Paternal haplotype

[0183] In another embodiment, methylation deconvolution analysis can be applied for the analysis of the paternal haplotype inheritance.

[0184] FIG. 12B shows an illustration of the paternal haplotype methylation analysis according to embodiments of the present invention. Methylation deconvolution can be performed on the maternal plasma DNA fragments carrying the alleles on paternal Hap III and Hap IV. As Hap III has been inherited by the fetus, the placental contribution would be higher for Hap III compared with Hap IV. Thus, the paternal inheritance of the fetus can be deduced.

[0185] This embodiment has advantages over previous methods based on the analysis of paternal-specific alleles. For example, for SNP at position 1, the A allele is present in the father, but not in the mother. Therefore, the detection of the paternal-specific A allele in maternal plasma indicates the inheritance of the Hap III by the fetus. However, for SNP at position 2, both the C and T alleles are not fetal-specific. In this situation, paternal-specific allele analysis cannot be used. However, the methylation deconvolution analysis does not require the presence of paternal-specific allele. Thus, SNPs that are heterozygous in both the father and the mother can be used for the methylation deconvolution analysis of the two paternal haplotypes.

[0186] Accordingly, a similar process, as used for maternal haplotypes, can be used to determine which paternal haplotype is inherited. In FIG. 12B, the placental contribution for Hap III would be higher than the placental contribution from Hap IV. The paternal haplotypes can be determined in a same or similar manner as the maternal haplotypes can be determined.

3. Method using Deconvolution

[0187] FIG. 13 is a flowchart illustrating a method 1300 of determining a portion of a fetal genome from a maternal sample using methylation deconvolution according to embodiments of the present invention. The biological sample includes a mixture of cell-free DNA molecules from a plurality of tissues types, including maternal tissue types and a fetal tissue type. The fetus has a father and a mother being the pregnant female. The portion of the fetal genome can be an entire chromosome copy or just part of the chromosome copy. The determined portions of the fetal genome can be combined to provide information on different portions of the fetal genome, up to the entire fetal genome.

[0188] At block 1310, a plurality of cell-free DNA molecules from the biological sample are analyzed. Block 1310 can be performed using techniques described in block 140 of method 100 of FIG. 1. For example, at least 1,000 cell-free DNA molecules can be analyzed to determine where

the cell-free DNA molecules are located, and methylation levels can be measured as described below. Further, the cell-free DNA molecules are analyzed to determine a respective allele of the cell-free DNA molecule. For example, an allele of a DNA molecule can be determined from a sequence read obtained from sequencing or from a particular probe that hybridizes to the DNA molecule, where both techniques can provide a sequence read (e.g., the probe can be treated as the sequence read when there is hybridization).

[0189] At block 1320, a first haplotype and a second haplotype of a first chromosomal region of a first parental genome of a first parent of the fetus are determined. One skilled in the art will be aware of various techniques to determine haplotypes of a parent. The haplotypes may be determined from the same sample as used to determine methylation levels below or from a different sample. In some implementations, the haplotypes can be determined from cellular samples, e.g., the buffy coat of a blood sample or the tissue of another organ. Examples of determining haplotypes are provided in U.S. Patent No. 8,467,976, which is incorporated by reference in its entirety. The first parent can be the mother or the father. Other examples of methods for detecting the parental haplotypes include, but not limited to the methods described by Fan et al (Nat Biotechnol 2011; 29: 51-57), Snyder et al (Nat Rev Genet 2015; 16: 344-358), the GemCode technology from 10X Genomics (www.10xgenomics.com/), and Targeted Locus Amplification (TLA) technology from Cergentis (www.cergentis.com/).

[0190] At block 1330, one or more heterozygous loci are identified from the first and second haplotypes. Each heterozygous locus has a corresponding first allele in the first haplotype and a corresponding second allele in the second haplotype. The one or more heterozygous loci may be a first plurality of heterozygous loci, where a second plurality of heterozygous loci can correspond to a different chromosomal region.

[0191] At block 1340, a first set of the plurality of cell-free DNA molecules is identified. Each of the plurality of cell-free DNA molecules is located at any one of the heterozygous loci from block 1330 and includes a corresponding first allele, so that the cell-free DNA molecule can be identified as corresponding to the first haplotype. It is possible for a cell-free DNA molecule to be located at more than one of the heterozygous loci, but typically a read would only include one heterozygous locus. Each of the first set of cell-free DNA molecules also includes at least one of N genomic sites, where the genomic sites are used to measure the methylation levels. N is an integer,

e.g., greater than or equal to 2, 3, 4, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000. Thus, a read of a cell-free DNA molecule can indicate coverage of 1 site, 2 sites, etc.

[0192] At block 1350, N first mixture methylation levels are measured at the N genomic sites (e.g., CpG sites) using the first set of the plurality of cell-free DNA molecules. One first mixture methylation level can be measured for each of the N genomic sites. Block 1350 can be performed in a similar manner as block 150 of method 100 of FIG. 1. In some embodiments, the measurement of the methylation level of a DNA molecule can use methylation-aware sequencing results, which may also be used to determine the location and respective allele of the DNA molecule. One skilled in the art will be aware of the various techniques that can be used to determine a methylation status of a site on a DNA molecule.

[0193] At block 1360, a first fractional contribution of the fetal tissue type in the mixture is determined using the N first methylation levels. In some embodiments, block 1360 can be performed via blocks 160 and 170 of method 100 of FIG. 1. Thus, a fractional contribution can be determined simultaneously for a panel of M tissue types. Block 1360 may use N issue-specific methylation levels at N genomic sites, determined for each of M tissue types, e.g., as in block 120 of method 100 of FIG. 1.

[0194] At block 1370, a second set of the plurality of cell-free DNA molecules is identified. Each of the plurality of cell-free DNA molecules is located at any one of the heterozygous loci from block 1330 and includes a corresponding second allele, so that the cell-free DNA molecule can be identified as corresponding to the second haplotype. Each of the second set of cell-free DNA molecules also includes at least one of the N genomic sites, where the genomic sites are used to measure the methylation levels.

[0195] At block 1380, N second mixture methylation levels at the N genomic sites are measured using the second set of the plurality of cell-free DNA molecules. Block 1380 may be performed in a similar manner as block 1350.

[0196] At block 1385, a second fractional contribution of the fetal tissue type in the mixture is determined using the N second methylation levels. Block 1385 may be performed in a similar manner as block 1360.

[0197] At block 1390, a first separation value is computed between the first fractional contribution and the second fractional contribution. Examples of separation values are described herein, e.g., including a difference or a ratio.

[0198] At block 1395, the portion of fetal genome is determined at the one or more heterozygous loci based on the first separation value. Thus, an inherited haplotype of the first parent can be determined. For example, the first separation value can be a ratio of the first fractional contribution and the second fractional contribution. The portion of the fetal genome can be determined to have one or more copies of the first haplotype and no copies of the second haplotype when the ratio is greater than a threshold value. Examples of threshold values include but not limited to 1.3, 1.4, 1.5, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8 and 3.0. The portion of the fetal genome can be determined to have one or more copies of the second haplotype and no copies of the first haplotype when the ratio is less than a threshold value. Examples of threshold values include but not limited to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8. The portion of the fetal genome can be determined to have the first haplotype and the second haplotype when the ratio is equal to one within a cutoff value. Examples of cutoff values include but not limited to 0.85, 0.9, 0.95, 1.0, 1.05, 1.1 and 1.15. Both haplotypes might be inherited when both parents have a same haplotype in the region being analyzed.

[0199] As another example, the first separation value is a difference of the first fractional contribution and the second fractional contribution. The portion of the fetal genome can be determined to have one or more copies of the first haplotype and no copies of the second haplotype when the difference is greater than a threshold value. Examples of threshold values include but not limited to 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 10%, 12%, 14%, 16%, 18% and 20%. The portion of the fetal genome can be determined to have one or more copies of the second haplotype and no copies of the first haplotype when the difference is less than a threshold value, e.g., where the threshold value is a negative number.

[0200] The inherited haplotype of the other parent can also be determined. For example, a second plurality of heterozygous loci of the first chromosomal region can be identified in the genome of the other parent. Fractional contributions can be determined for each of the other parent's haplotypes, and a separation value can be used to determine the inherited haplotype of the other parent.

[0201] For example, the first plurality of heterozygous loci and the second plurality of heterozygous loci can be the same loci or be different. Each of the second plurality of heterozygous loci can include a corresponding third allele in a first haplotype of the other parent (e.g., a first paternal haplotype) and a corresponding fourth allele in a second haplotype of the other parent (e.g., a second paternal haplotype). The third and fourth alleles can be the same as the first and second alleles. In addition to the first and second set of cell-free DNA molecules for the first parent, a third set of the plurality of cell-free DNA molecules can each be located at any one of the second plurality of heterozygous loci, include the corresponding third allele of the heterozygous locus, and include at least one of K genomic sites. The K genomic sites may be the same or different than the N genomic sites used for the first parent. In a similar manner as with the first parent, K third mixture methylation levels can be measured at the K genomic sites using the third set of the second plurality of cell-free DNA molecules, and a third fractional contribution of the fetal tissue type in the mixture can be determined using the K third methylation levels. The third fractional contribution corresponds to the first haplotype of the other parent (e.g., the first paternal haplotype).

[0202] A fourth set of the plurality of cell-free DNA molecules can each be located at any one of the second plurality of heterozygous loci, include the corresponding fourth allele of the heterozygous locus, and include at least one of the K genomic sites. Thus, the fourth set of DNA can be used to test the second haplotype of the other parent. K fourth mixture methylation levels at the K genomic sites can be measured using the fourth set of the second plurality of cell-free DNA molecules, and a fourth fractional contribution of the fetal tissue type in the mixture can be determined using the K fourth methylation levels. A second separation value can be computed between the third fractional contribution and the fourth fractional contribution, and the portion of fetal genome at the second plurality of heterozygous loci can be determined based on the second separation value. The inherited haplotype from the other parent can be determined in a similar manner as for the first parent. The fourth fractional contribution corresponds to the second haplotype of the other parent (e.g., the second paternal haplotype).

[0203] In some embodiments, the second fractional contribution does not need to be determined. Instead, a haplotype can be determined to be inherited if the corresponding fractional contribution is sufficiently high. For example, the first fractional contribution can be compared to a reference

value to determine whether the fetus inherited the first haplotype at the first chromosomal region. The fetus can be determined to have inherited the first haplotype at the first chromosomal region when the first fractional contribution exceeds the reference value.

[0204] In other embodiments, the reference value may be determined from the second fractional contribution. For example, the reference value can be a sum of the second fractional contribution and a threshold value. The sum with the threshold value can ensure that the first fractional contribution is sufficiently greater than the second fractional contribution.

[0205] A separate determination of inheritance can be made for the second haplotype by comparing the second fractional contribution to the reference value to determine whether the fetus inherited the second haplotype at the first chromosomal region. The fetus can be determined to have inherited the second haplotype at the first chromosomal region when the second fractional contribution exceeds the reference value. If both the fractional contributions are determined to exceed the reference value, the two fractional contributions can be compared to each other to determine if one is significantly greater than the other (e.g., using a threshold). The haplotypes of the other parent can be determined to identify whether one of these haplotypes is the same as the haplotypes of the first parent, thereby explaining that both haplotypes of the first parent could have been inherited.

C. Determination of Inherited Haplotype using Methylation Levels

[0206] Other embodiments can use the general hypomethylation of cell-free fetal DNA to identify the inherited haplotype as the one with the lower overall methylation level. Embodiments can start with deduced maternal or paternal haplotypes, and then measure the methylation levels of plasma DNA molecules containing SNP alleles in each of those deduced haplotypes. In one implementation of analyzing the maternal haplotypes, the methylation levels of the two deduced maternal haplotypes can be compared, and the one with the lower methylation level would be predicted to be the haplotype inherited by the fetus. In another implementation of analyzing the paternal haplotypes, the methylation levels of the two deduced paternal haplotypes can be compared, and the one with the lower methylation level would be predicted to be the haplotype inherited by the fetus.

1. Example

[0207] As an example, a methylation level of each of the two maternal haplotypes can be determined. As the placental tissue is relatively hypomethylated compared with other tissues, it is expected that the maternal haplotype inherited by the fetus would be more hypomethylated than the one that is not inherited by the fetus. The methylation densities were tested in the maternal plasma using the actual haplotypes of the mother, which were deduced using the maternal, paternal, and fetal genotypes.

	Hap I	Hap II
Overall methylation densities	65%	87%

Table 9. Methylation densities for actual Hap I and Hap II.

[0208] Table 9 shows methylated densities of the two maternal haplotypes in the maternal plasma. As Hap I was the actual haplotype inherited by the fetus by genotyping, the results of the methylation analysis of the haplotype correctly identified the inheritance.

[0209] In other embodiments, the maternal haplotypes can be deduced based on the genotypes of the mother alone, or reference haplotypes of the population from haplotype database can also be used for this analysis. The maternal haplotypes used in this example were phased using the IMPUTE2 program. Thus, deduced maternal haplotypes can also be used in this analysis.

	Hap I	Hap II
Overall methylation densities	68%	76%

Table 10. Methylation densities for deduced Hap I and Hap II.

[0210] Table 10 shows methylated densities of the two deduced maternal haplotypes in the maternal plasma. The deduced maternal haplotype that was inherited by the fetus had lower methylation densities. An example of a statistical procedure that one can use to determine whether one haplotype has a sufficiently lower methylation density includes the chi-square test. A separation between the two methylation levels can be required to be sufficiently large (e.g., greater than a threshold) to make the determination. If the separation is not sufficient, then an indeterminate classification can be made. In some embodiments, a determination of an inheritance of both haplotypes can be determined if the separation is not sufficiently large and if both methylation levels are below a threshold level, which may be characterized by inclusion of fetal

DNA. For example, tables 9 and 10 indicate that a methylation density below 70% may indicate that the fetus has inherited that haplotype. Both haplotypes may be inherited when the parents share a haplotype for the region being analyzed.

[0211] In another embodiment, the overall methylation densities of the maternal plasma DNA carrying the paternal Hap III and Hap IV can be compared. Similar to the maternal haplotype analysis, the fetus would be deduced as having inherited the paternal haplotype that has lower overall methylation densities.

2. Method using Methylation Levels

[0212] FIG. 14 is a flowchart illustrating a method 1400 of determining a portion of a fetal genome from a maternal sample using methylation levels according to embodiments of the present invention. The biological sample includes a mixture of cell-free DNA molecules from a plurality of tissues types, including maternal tissue types and a fetal tissue type. The fetus has a father and a mother being the pregnant female. The portion of the fetal genome can be an entire chromosome copy or just part of the chromosome copy. The determined portions of the fetal genome can be combined to provide the entire fetal genome, as with other methods described herein.

[0213] At block 1410, a plurality of cell-free DNA molecules from the biological sample are analyzed. Block 1410 can be performed in a similar manner as block 1310 of method 1300 of FIG. 13.

[0214] At block 1420, a first haplotype and a second haplotype of a first chromosomal region are determined of a first parental genome of a first parent of the fetus. Block 1420 can be performed in a similar manner as block 1320 of FIG. 13. In some embodiments, genotypes of the first parent's genome can be determined at the plurality of heterozygous loci using a sample from the first parent, e.g., a blood sample or other tissue that may or may not include fetal DNA. A plurality of reference haplotypes can be obtained, e.g., from databases of reference genomes. The first haplotype and the second haplotype can be deduced using the genotypes and the plurality of reference haplotypes. For example, the alleles of each genotype can be compared against the reference haplotypes, and any haplotypes that do not include the alleles at the corresponding locus can be discarded. Once two reference haplotypes remain, those haplotypes can be identified as being the first haplotype and the second haplotype.

[0215] At block 1430, a plurality of heterozygous loci is identified from the first and second haplotypes. Each heterozygous locus has a first allele in the first haplotype and a second allele in the second haplotype.

[0216] At block 1440, a first set of the plurality of cell-free DNA molecules is identified. Block 1440 can be performed in a similar manner as block 1340 of FIG. 13.

[0217] At block 1450, a first mixture methylation level is measured using the first set of the plurality of cell-free DNA molecules. For example, the first mixture methylation level may be a methylation density for the cell-free DNA molecules of the first set. The methylation density can be computed as a total methylation density for all of the cell-free DNA molecules of the first set. In another example, separate methylation densities can be computed for each locus, and the separate methylation densities can be combined to obtain the first mixture methylation level, e.g., an average of the separate methylation densities.

[0218] At block 1460, a second set of the plurality of cell-free DNA molecules is identified. 1460 can be performed in a similar manner as block 1370 of FIG. 13.

[0219] At block 1470, a second mixture methylation level is measured using the second set of the plurality of cell-free DNA molecules. For example, the second mixture methylation level may be a methylation density for the cell-free DNA molecules of the second set.

[0220] At block 1480, it is determined which of the first haplotype and the second haplotype is inherited by the fetus based on which of the first mixture methylation level and the second mixture methylation level is lower. As part of block 1480, a separation value can be determined between the first mixture methylation level and the second mixture methylation level, and compared to a threshold value. The threshold value can ensure that the lower level is sufficiently lower. The threshold value can be determined using the chi-square test. For example, measurements can be taken of samples where the inherited haplotype is known, and a distribution of the separation values can be determined, and a threshold value can be selected that accurately determines the inherited haplotype in the training data obtained from the samples. Methods 1300 and 1400 can also be combined, with each being performed as a check, and the inherited haplotype determined if both methods are consistent with each other.

D. Selection of loci

[0221] Various embodiments can be used for the comparison of the methylation levels or the fraction contributions of the two deduced maternal haplotypes in the maternal plasma. In one embodiment, the number of SNP loci to be analyzed can be determined before the analysis. For example, the number of SNP loci used in the haplotype deconvolution analysis can be determined according to a number of factors, for example, but not limited to the desired statistical power, the mean difference in methylation levels in the placenta and blood cells in the region of interest, and the number of molecules being analyzed for each SNP.

[0222] The size of the region of interest can be fixed, and all SNPs within the region of interest can be used in the analysis. The size of the region of interest can be determined taking into account a number of factors, for example, but not limited to the desired statistical power, the mean difference in methylation level in the placenta and blood cells in the region of interest, the number of molecules being analyzed for each SNP, and the chance of meiotic recombination with the region of interest.

[0223] In other embodiments, the number of SNPs and the size of the region to be analyzed are not determined before the analysis. For example, the number of SNPs can be sequentially increased until the data are sufficient to arrive at a statistically significant conclusion concerning which maternal haplotype is statistically significantly less methylated than the other one. For instance, the SNPs on the region of interest can be arranged in an ascending order of their genomic coordinates. Then statistical testing can be carried out with the data of the SNP with the lowest number of genomic coordinate. If this is sufficient to make a conclusion regarding which haplotype is less methylated statistically, then a conclusion is made. Similarly, the SNPs can be arranged in a descending order, with a highest number of genomic coordinate that is sufficient being used.

[0224] If the statistical accuracy is not sufficient, another statistical comparison can be performed starting from the next SNP with a higher number of genomic coordinates. On the other hand, if the data of the first SNP are not sufficient for one to conclude that one haplotype is less methylated than the other one (or that the separation value between fractional contributions is not sufficiently large), the data of another SNP can be added and another round of statistical testing is

carried out. This procedure can be continued until the accumulated data are sufficient to make a statistically significant conclusion. A number of statistical tests can be performed to compare the methylation levels of the two haplotypes, for example, but not limited to Student's t-test, Mann-Whitney rank-sum test and Chi-square test. The level of statistical significance can be determined based on the desired confidence of the conclusion, for example, but not limited to adopting a P-value of 0.05, 0.01, 0.001, 0.0001, or 0.00001.

E. Combinations with RHDO

[0225] In some embodiments, results generated by the RHDO analysis of U.S. Patent 8,467,976 can be combined with present methylation embodiments to arrive at a more accurate procedure for diagnosis or to reduce an amount of sequencing required. For example, fetal haplotypes can be determined using present embodiments and using results of the RHDO analysis of U.S. Patent 8,467,976, and the determined fetal haplotypes from both of the techniques can be compared. For example, the results from the two analyses would be accepted only if they are concordant. Further analysis can be performed if the two analyses show different conclusions, e.g., measurements can be repeated at higher depth of coverage on the genome.

[0226] For such a combined approach to be most cost-effective, it is preferred to have one type of sequencing that can yield data for both methods. In one embodiment, this can be done by a single molecule method that would generate sequencing as well as methylation information, e.g. using the Single Molecule Real Time sequencing technology from Pacific Biosciences, or nanopore sequencing (e.g. from Oxford Nanopore Technologies). These are two examples of methylation-aware sequencing. In another embodiment, the RHDO analysis can be performed on bisulfite sequencing results. For such an embodiment, any maternal and paternal genetic information can be determined also using bisulfite sequencing. Bisulfite sequencing is thus another example of methylation-aware sequencing. Furthermore, other methylation aware sequencing technologies can be used, such as oxidative bisulfite sequencing (Booth et al. Science 2012; 336: 934-937) or Tet-assisted bisulfite sequencing (Yu et al. Cell 2012; 149: 1368-1380). The latter examples would allow one to analyze the 5-methylcytosine distribution of the analyzed DNA molecules.

F. Uses of knowledge of fetal genome

[0227] The noninvasive prenatal analysis of the fetal genome can be used to determine if a fetus has inherited a disease from the parents. This is particularly useful for the detection of monogenic diseases, for example congenital adrenal hyperplasia (New et al. J Clin Endocrinol Metab 2014;99:E1022-30), beta-thalassemia (Lam et al. Clin Chem. 2012;58:1467-75) and hereditary muscular dystrophies (Genet Med 2015;17:889-96). If a monogenic disease is detected, various treatments can be performed, e.g., the pregnancy can be terminated, treatment provided before pregnancy, or after birth. For example, steroid treatment can be given prenatally to a pregnant woman confirmed of having a fetus affected by congenital adrenal hyperplasia to avoid abnormal sexual development.

VI. HAPLOTYPE DECONVOLUTION ANALYSIS FOR ANEUPLOIDY DETECTION

[0228] The haplotype deconvolution can also be used to detect a sequence imbalance of a chromosomal region of a fetus, such as aneuploidies, microdeletions, or microamplifications (e.g., microduplications). For example, a fractional contribution of a haplotype in one region can be compared to a fractional contribution of another haplotype in another region.

A. Mother

[0229] FIG. 15 shows a chromosomal aneuploidy detection based on haplotype deconvolution for maternal haplotypes according to embodiments of the present invention. In this illustration, the mother has two maternal haplotypes, namely Hap I and Hap II. For illustration purposes, we assume that 80% of her plasma DNA was derived from her own cells and 20% was derived from the placenta, which are example percentages in commonly measured ranges. This method can be generally applied for pregnancies with different fetal DNA percentages. Knowledge of the fetal DNA percentage is not required, but is simply provided for illustration, although a measurement of the fetal DNA percentage may be performed in various ways, e.g., using fetal-specific alleles or fetal-specific methylation markers.

[0230] The fetus has inherited Hap I and another haplotype from the father, namely Hap III. The placentally-derived DNA would exhibit the fetal genotypes, and a sequence imbalance can be detected by analyzing the fractional contribution resulting from the placentally-derived DNA.

[0231] As illustrated above, the fetal inheritance of the maternal haplotype can be determined through the deconvolution of the two maternal haplotypes. Analysis for the placental contribution to the maternal DNA can be performed for each of the two maternal haplotypes. The maternal haplotype inherited by the fetus (Hap I in this example) would have a much higher placental contribution compared with the maternal haplotype that is not inherited by the fetus (Hap II). The placental contribution for Hap I would be positively correlated with the fetal DNA fraction in the maternal plasma.

[0232] After determining which maternal haplotype has been inherited by the fetus, the dosage of the chromosome the fetus has inherited from the mother can be further determined through maternal haplotype deconvolution. In this illustration, two chromosomal regions are analyzed using maternal haplotype deconvolution. In one embodiment, the reference chromosome (RefChr) is a chromosome or a chromosomal region which is unlikely to be affected by a chromosomal aneuploidy. The reference chromosomal region is shown on the left side of FIG. 15. The target chromosome (TargetChr) is a chromosome or a chromosomal region which is potentially affected by a chromosomal aneuploidy. The target chromosomal region is shown on the right side of FIG. 15. The two regions can be for different regions of a same chromosome or for regions of two different chromosomes.

[0233] In the example shown, the fetus has been deduced to have inherited Hap I from the mother for both the reference chromosome and target chromosome through the methylation deconvolution of the Hap I and the Hap II at each region. Then, the placental contribution to maternal plasma DNA for Hap I can be compared between the reference chromosome and the target chromosome. If the placental contribution of Hap I for the target chromosomal region is significantly different than the placental contribution of Hap I for the reference chromosomal region (e.g., higher for amplification or lower for deletion), then a sequence imbalance can be identified.

[0234] For illustration purpose, we use the detection of trisomy as an example. However, other types of chromosome aneuploidies, including monosomy, amplification of a subchromosomal region or deletion of a subchromosomal region can also be detected using this method. For trisomy, the extra copy of the affected chromosome can be inherited from the father (denoted as Trisomy (F)) or the mother (denoted as Trisomy (M)). In over 90% of trisomy 21 cases, the extra

copy of chromosome 21 is derived from the mother (Driscoll et al. N Engl J Med 2009;360: 2556–2562). In the scenario of Trisomy (M), the placental contribution of Hap I for the target chromosome would be higher than that for the reference chromosome. In FIG. 15, Trisomy (M) is shown with two instances of Hap I, which would provide a higher placental contribution for the target region than the one instance of Hap I for the reference region.

[0235] Whether the placental contribution of Hap I for the target chromosome is higher than that for the reference chromosome can be determined by comparing a separation value between the two placental contributions and a threshold, which may be based on a separate measurement of the fetal DNA percentage. A higher fetal DNA percentage would result in a higher expected separation value between the two placental contributions, and thus the threshold can be set higher. For example, with the fetal DNA percentage being 20%, the placental contribution of Hap I for the reference region would be about 20% and the placental contribution of Hap I for the target region would be about 36.4%.

[0236] For instance, assume that 10 DNA molecules exist at the reference chromosome, then two of them are fetal and eight of them are maternal. For the two fetal DNA molecules, one is derived from Hap I and one is derived from Hap III. For the eight maternal DNA molecules, four are Hap I and four are Hap II. For the target region, there would be an extra DNA molecule of Hap I from the fetus. Thus there would be two fetal Hap I DNA molecules and 4 maternal Hap I DNA molecules total, providing 2/6=33.3%. The threshold value for the difference (e.g., 13.3%) can be placed between 0 and 13.3% to provide optimal specificity and sensitivity. A distribution of separation values can be determined from a reference group of samples. In the scenario of Euploid, the placental contributions would be approximately equal, e.g., the separation value would be less than the threshold. One skilled in the art will know how to select a suitable threshold based on the description herein and in U.S. Patent 8,467,976, and other references cited herein.

[0237] In one embodiment, the ratio (or other separation value) of placental contribution of Hap I between the target and reference chromosomes for a group of pregnant women, each known to be carrying a euploid fetus can be used as a reference interval. The ratio in the tested case can be compared with this reference group to determine if a significant elevation of placental contribution of Hap I is present for the target region relative to the reference region. In the example of 20% fetal DNA, the ratio would be 33.3/20 = 1.67. The ratio can be generalized to 2/(1+f), where f

represents the fetal DNA fraction. In another embodiment, the difference in the placental contribution of Hap I between the target and the reference chromosomes can be determined. This difference is then compared with a reference group.

B. Father

[0238] In another embodiment, haplotype deconvolution of the paternal haplotypes (Hap III and Hap IV) can be performed in the maternal plasma. The analysis of the paternal haplotypes can be performed in a similar manner as for the maternal haplotypes.

[0239] FIG. 16 shows a chromosomal aneuploidy detection based on haplotype deconvolution for paternal haplotypes according to embodiments of the present invention. In this illustration, the father has two paternal haplotypes, namely Hap III and Hap IV. As in FIG. 15, the fetus has inherited Hap I from the mother and Hap III from the father.

[0240] In the scenario where the extra copy of chromosome is derived from the father (Trisomy (F)), the placental contribution of Hap III would be higher for the target chromosome than for the reference chromosome. This is shown for the Trisomy (F) example, where two copies of Hap III are shown. As described above for the maternal haplotypes, a separation value between the placental contributions of Hap III for the target and reference regions can be compared to a threshold to determine whether an extra copy of Hap III exists for the target region. In various embodiments, a ratio or difference of the two placental contributions of the tested case can be compared with a reference group of pregnant women, each known to be carrying a euploid fetus so as to determine if the fetus has a chromosomal trisomy for the target chromosome, or amplification or deletion of a target chromosomal region. The threshold can be based on the separation values for the reference group of euploid fetuses, a reference group of aneuploidy fetuses, or both. A separate measure of fetal DNA percentage can also be used, as is described herein.

C. Method of detecting a sequence imbalance

[0241] FIG. 17 is a flowchart of a method 1700 for detecting a sequence imbalance in a portion of a fetal genome of an unborn fetus of a pregnant female using a biological sample from the pregnant female according to embodiments of the present invention.

[0242] At block 1710, a plurality of cell-free DNA molecules from the biological sample are analyzed. Block 1710 can be performed in a similar manner as block 1310 of method 1300 of FIG. 13.

[0243] At block 1720, a first target haplotype of a target chromosomal region of a first parental genome of a first parent of the fetus is determined, and a first reference haplotype of a reference chromosomal region of the first parental genome is determined. Block 1720 can be performed in a similar manner as block 1320 of FIG. 13. The target chromosomal region and the reference chromosomal region can be an entire chromosome or just part of a chromosome. Thus, the target chromosomal region can be a first chromosome and the reference chromosomal region can be a second chromosome different from the first chromosome. The first parent can be the mother or the father of the fetus.

[0244] The target chromosomal region can be selected based on various criteria. For example, a plurality of target regions can be selected, as may occur to test many non-overlapping regions of a specified size, such as 1 Mb, 5 Mb, 10 Mb, 20 Mb, 50 Mb, etc. As another example, the target chromosomal region can be selected based on a copy number analysis that identifies the region as having more DNA molecules than expected, e.g., as described in U.S. Patent Publications 2009/0029377 and 2011/0276277.

[0245] In some embodiments, it can be determined that the fetus has inherited the first target haplotype from the first parent and that the fetus has inherited the first reference haplotype from the first parent. The determination can include embodiments of FIG. 13 or FIG. 14. For example, determining that the fetus has inherited the first target haplotype from the first parent can include determining a second target fractional contribution of the fetal tissue type in the mixture corresponding to the second target haplotype, computing a second separation value between the first target fractional contribution and the second target fractional contribution, and determining that the fetus has inherited the first target haplotype from the first parent based on the second separation value.

[0246] At block 1730, a plurality of target heterozygous loci are identified of the target chromosomal region of the first parental genome. Each target heterozygous locus includes a corresponding first target allele in the first target haplotype and a corresponding second target allele in a second target haplotype of the first chromosomal region of the first parental genome.

Referring back to the example of FIG. 15, the target heterozygous loci have corresponding first target alleles of {G,T,A} on Hap I and have corresponding second target alleles of {A,G,C} on Hap II.

[0247] At block 1740, a target set of the plurality of cell-free DNA molecules is identified. Each cell-free DNA molecule of the target set is located at any one of the target heterozygous loci, includes a corresponding first target allele, and includes at least one of N genomic sites in the target chromosomal region. Block 1740 can be performed in a similar manner as described herein. For example, sequence reads can be mapped to a reference genome, where the target set of plurality of cell-free DNA molecules aligns to any one of the target heterozygous loci.

[0248] At block 1750, N first mixture methylation levels are measured at the N genomic sites using the target set of the plurality of cell-free DNA molecules. Block 1750 can be performed in a similar manner as block 1350 of FIG. 13.

[0249] At block 1760, a first fractional contribution of the fetal tissue type in the mixture is determined using the N first methylation levels. Block 1760 can be performed in a similar manner as block 1360 of FIG. 13.

[0250] At block 1770, a plurality of reference heterozygous loci are identified for the reference chromosomal region of the first parental genome. Each reference heterozygous locus includes a corresponding first reference allele in the first reference haplotype and a corresponding second reference allele in a second reference haplotype of the reference chromosomal region of the first parental genome. Referring back to the example of FIG. 15, the reference heterozygous loci have corresponding first target alleles of {A,T,C} on Hap I and have corresponding second target alleles of {T,C,A} on Hap II.

[0251] At block 1775, a reference set of the plurality of cell-free DNA molecules is identified. Each cell-free DNA molecule of the reference set is located at any one of the reference heterozygous loci, includes a corresponding first reference allele, and includes at least one of K genomic sites in the reference chromosomal region.

[0252] At block 1780, K reference mixture methylation levels are measured at the K genomic sites using the reference set of the plurality of cell-free DNA molecules.

[0253] At block 1785, a first reference fractional contribution of the fetal tissue type is determined in the mixture using the K reference methylation levels.

[0254] At block 1790, a first separation value is computed between the first target fractional contribution and the first reference fractional contribution.

[0255] At block 1795, the first separation value is compared to a threshold value to determine a classification of whether the fetus has a sequence imbalance for the target chromosomal region. If the first separation value exceeds the threshold value then a sequence imbalance can be identified. The threshold value can be determined as described above, e.g., based on separation values seen in a reference group of samples not having a sequence imbalance and/or a reference group of samples having the sequence imbalance. As examples, the classification can be positive, negative, or indeterminate for the sequence imbalance being tested.

[0256] Different threshold values can be used, depending on the type of sequence imbalance. For example, if the sequence imbalance is a deletion, then the first separation value would be expected to be a negative value. In such a case, the threshold value can be a negative number, and the comparison can determine that the first threshold value exceeds the threshold value by being a larger negative number. If the sequence imbalance being tested is an amplification, then it can be tested whether the separation value is greater than the threshold value. Thus, the threshold value used can be dependent on the type of sequence imbalance being tested.

VII. DECONVOLUTION OF SIGNATURES TO IDENTIFY DISEASED TISSUE

[0257] If a genomic signature (e.g., a particular SNP allele) is known, embodiments can determine which tissue is the origin of such signatures. As the cell-free DNA molecules exhibiting the signatures are from the tissue of origin, the tissue of origin can be identified from the fractional contributions determined using cell-free DNA molecules exhibiting the signatures. Thus, cell-free DNA molecules with a signature of a transplanted organ (e.g., a signature of a haplotype of the transplanted organ) can be used to monitor changes in amounts of cell-free DNA molecules from the transplanted organ with high sensitivity, e.g., given that a high fractional contribution of the DNA in the mixture would be from the transplanted organ. Examples are provided for transplants to show that the technique is accurate. In another example, a signature of a tumor can be used to identify tissues within which the tumor resides.

A. Organ Transplantation

[0258] As an example for organ transplantation, we analyzed the plasma of a patient who had received liver transplantation and a patient who had received bone marrow transplantation. For each case, the donor-specific SNP alleles were identified through the genotyping of the tissues from the patients and the donors. For the liver transplant recipient, a biopsy of the donor liver and the blood cells of the recipient were sequenced. For the bone marrow transplant case, the buccal swab (recipient genotype) and the blood cells (donor genotype) were sequenced. The plasma DNA samples were sequenced after bisulfite conversion. Sequenced DNA fragments carrying a donor-specific SNP allele and at least one CpG site were used for downstream methylation deconvolution analysis. A total of 72 million and 121 million reads were sequenced for the patients who had received liver and bone marrow transplantation, respectively. For the two cases, 38 and 5355 fragments were used for deconvolution analysis, respectively.

Tissue type	Liver transplant recipient	Bone marrow transplant recipient
Liver	45.4	4.4
Lung	0.0	1.5
Colon	29.3	6.3
Small intestines	0.0	1.8
Pancreas	0.0	0.0
Adrenal glands	0.0	0.0
Esophagus	0.0	0.0
Adipose tissues	0.0	14.8
Heart	0.0	0.0
Brain	14.5	9.6
T-cells	0.0	12.3
B-cells	5.9	16.6
Neutrophils	4.9	32.8

Table 6. Fractional contributions for different organs to plasma DNA fragments carrying donor-specific alleles in the two transplant recipients.

[0259] Table 6 shows methylation deconvolution analysis on plasma DNA fragments carrying donor-specific alleles in a liver transplant recipient and a bone marrow transplant recipient. The

numbers represent the percentage contribution of different tissues to the donor-specific plasma DNA fragments. For the liver transplant case, the liver was shown to be the most important contributor to these DNA fragments. For the bone marrow transplant case, the hematopoietic system (including the T-cells, B-cells and neutrophils) was the major contribution of the donor-specific DNA fragments. These results indicate that methylation deconvolution can accurately indicate the tissue origin of DNA fragments having single nucleotide alterations. A small amount of sequenced fragments were attributed to other tissues probably because of measurement imprecision as relatively small number of donor-specific fragments were used for the deconvolution analysis.

[0260] The fractional contributions for tissue associated with the transplanted organ can be determined in the above manner and monitored. With the baseline fractional contribution (an example of a reference fractional contribution) relatively high as a result of using only cell-free DNA molecules exhibiting the donor signature, small changes in total amount of donor DNA in plasma can be detected. Accordingly, methylation deconvolution analysis can be applied for the monitoring of organ transplantation.

[0261] As can be seen above for the liver transplant, methylation deconvolution is not absolutely specific. In this analysis, plasma DNA fragments carrying the donor-specific alleles were used for methylation deconvolution analysis. These fragments are specific for the donor and should be derived only from the liver in this liver transplant recipient. Therefore, the theoretical contribution of the liver should be 100%. Another possibility is that certain cell types are present in different types of tissues making the liver methylation profile overlapping with other tissues. For example, the connective tissue cells in the liver may also be present in other organs. But, relative percentages from other patients or other samples (e.g., at other times) of the instant patient can identify whether more cell-free DNA molecules are being released.

[0262] In various embodiments, the donor signature can correspond to a particular haplotype of the donor genome or both haplotypes in a chromosomal region. Methylation deconvolution can be performed using cell-free DNA molecules located on the particular donor haplotypes, and increases in the fractional contribution of the particular haplotype can be monitored. If a significant increase occurs (e.g., as measured by a percentage or absolute threshold), then a rejection of the transplanted organ can be identified.

[0263] FIG. 18 shows an illustration of haplotype deconvolution for organ transplant monitoring according to embodiments of the present invention. The donor has haplotypes labeled Hap I and Hap II, and the recipient has haplotypes labeled Hap III and Hap IV. The donor has a signature at locus 1 and locus 3, as the alleles are not found on the recipient haplotypes. Locus 2 and locus 4 do not have a donor signature. Thus, embodiments may use DNA molecules that are located at locus 1 and locus 3 as part of a deconvolution process.

[0264] The plasma DNA deconvolution can be used to determine whether the determined fractional contribution from the transplanted organ is at a baseline or increased relative to the baseline. In some embodiments, the fractional contributions can be determined for each of Hap I and Hap II separately, if different signatures exist; such different signatures can exist at different loci. In other embodiments, a single fractional contribution can be determined for both haplotypes, e.g., when they share a signature. In the example shown in FIG. 18, Hap I and Hap II do share a signature at locus 1 and locus 3.

[0265] Accordingly, the contribution of the transplanted organ can be determined using haplotype deconvolution. The increase in the contribution of the haplotype to the transplanted organ would be useful to indicate the increased contribution of the organ to the plasma DNA. In various embodiments, the baseline level can be determined from a cohort of transplant recipients not having rejection or from a cohort of transplant recipients having rejection. When using recipients having rejection, the baseline level can be determined as below those from a cohort of transplant recipients having rejection.

[0266] As mentioned above, the donor may have two identical haplotypes or the recipient can also have two identical haplotypes. Furthermore, the donor and recipient may share a haplotype. As long as the donor or the recipient has a unique haplotype, a change in a percentage of cell-free DNA molecules from the donor tissue can be determined. In the former, a rejection will be detected when one sees an increase in the contribution of the donor-unique haplotype in plasma (or other sample). In the latter, a rejection will be detected when one sees a decrease in the contribution of the recipient-unique haplotype in plasma.

[0267] Accordingly, some embodiments can use a first haplotype that is present in normal cells of the organism and not being present in abnormal cells that may be in the mixture. This would correspond to the latter example above, when the recipient has a unique haplotype. Another

example is when a patient has a unique haplotype in healthy cells compared to a tumor (e.g., previously found in the organism). In this embodiment, the first tissue type can be determined to have the disease state when the first separation value is less than the threshold value.

[0268] In some embodiments, if the transplanted organ is detected as being rejected, treatment can be provided. For example, a change is dosage of anti-rejection medication can be provided. As another example, a new organ can be obtained, and surgery can be performed to remove the old transplanted organ and put in the new transplanted organ.

B. Hepatocellular carcinoma (HCC)

[0269] As an example for determining a tissue of origin for a cancer signature or aberration (or monitoring for a tumor that was known to exist or have existed), we analyzed the plasma of an HCC patient. The tumor and the blood cells of the patients were sequenced to identify the cancer-specific single nucleotide mutations. Sequenced DNA fragments carrying a cancer-specific mutation and at least one CpG site were used for downstream methylation deconvolution analysis. A total of 11,968 fragments were used for deconvolution analysis. In addition to the methylation profiles from the normal tissue organs, we have also included the methylation profile of HCC tissues as candidate tissue of origin.

[0270] In another embodiment, more types of tumor tissues can be considered as candidate tissues for the mutations. In one embodiment, the methylation profiles of the common cancers, for example but not limited to colorectal cancer, lung cancer, breast cancer, pancreatic cancer, prostate cancer, bladder cancer, cervical cancer and ovarian cancer can be included as candidate tissues. In yet another embodiment, only the most possible cancers specific to the patient can be included in the analysis. For example, in female patients, the breast cancer, ovarian cancer, colorectal cancer and cervical cancer are considered. In yet another embodiment, the ethnic origin and the age are considered in the selection of the candidate tissues.

[0271] Table 7 shows methylation deconvolution of the plasma DNA fragments carrying cancer-associated mutations. The deconvolution analysis accurately determined that the DNA fragments carrying cancer-associated mutations are predominantly derived from the liver cancer tissues.

Tissue	Contribution (%)

Liver	0.0
Lung	0.0
Colon	0.0
Small intestines	0.0
Pancreas	0.0
Adrenal glands	0.0
Esophagus	0.0
Adipose tissues	0.0
Heart	0.0
Brain	0.0
T-cells	0.0
B-cells	0.0
Neutrophil	4.6
Liver cancer	95.4
Placenta	0.0

Table 7. Fractional contributions for HCC patient using cancer mutation.

[0272] In some embodiments, the tumor can initially be identified by detecting a copy number aberration, e.g., as described in U.S. Patent Nos. 8,741,811 and 9,121,069. The particular tissue of origin can be determined, e.g., as described in U.S. Patent Application 14/994,053 based on patterns of copy number aberrations previously identified in various tumors. Once the tumor has been identified, treatment can be performed, e.g., by surgery, radiotherapy, or chemotherapy. Either way, a biopsy can be obtained after the tissue of origin is determined. A cancer-specific point mutation can be determined from the biopsy or from DNA fragments in plasma (e.g., as described in U.S. patent Publication 2014/0100121, or other mixtures that are associated with the copy number aberration.

[0273] After treatment, a key change would be the disappearance of the genomic aberrations, including the copy number aberration and point mutation. When these aberrations are gone, the analysis of the genomic signature of the point mutation in the affected regions would give a change in the tissue contribution via the methylation deconvolution analysis. If the tumor comes back in the future, the cancer-associated changes in tissue composition (as determined using methylation deconvolution analysis) would be seen again. For example, the fractional contribution can be compared to a reference fractional contribution, and if a change is detected, then new courses of treatment can be provided.

[0274] In various embodiments, the cancer-specific mutation can be on only one haplotype or on both haplotypes, e.g., in a manner similar to the donor example above. Thus, as with the donor, the fractional contributions can be determined for each of Hap I and Hap II separately, if different signatures exist; such different signatures can exist at different loci. In other embodiments, a single fractional contribution can be determined for both haplotypes, e.g., when they share a signature.

C. Imprinting

[0275] In another embodiment, the haplotype deconvolution analysis can be applied for the analysis of the genomic regions showing tissue-specific imprinting. It has been shown that the differential methylation of the paternally and maternally inherited alleles in different tissue organs is a common phenomenon (Baran et al. Genome Res 2015;25:927-36). Haplotype deconvolution would be useful for the monitoring of the contribution of the organ exhibiting tissue-specific imprinting. For example, when the paternally and maternally inherited haplotypes have different methylation status in the liver but not in other tissues, methylation deconvolution can be performed on both the paternally and maternally inherited haplotypes. In one embodiment, both the paternal and maternal methylation patterns can be included as candidate tissues in the analysis.

D. Method using Genomic Signature

[0276] FIG. 19 is a flowchart illustrating a method 1900 of analyzing a biological sample of an organism to detect whether a first tissue type has a disease state associated with a first haplotype according to embodiments of the present invention. The biological sample includes a mixture of cell-free DNA molecules from a plurality of tissues types, including a first tissue type. Method 1900 is performed at least partially using a computer system.

[0277] At block 1910, a plurality of cell-free DNA molecules from the biological sample are analyzed. Block 1910 can be performed using techniques described in block 140 of method 100 of FIG. 1. For example, at least 1,000 cell-free DNA molecules can be analyzed to determine where the cell-free DNA molecules are located, and methylation levels can be measured as described below. Further, the cell-free DNA molecules are analyzed to determine a respective allele of the cell-free DNA molecule. For example, an allele of a DNA molecule can be determined from a sequence read or from a particular probe that hybridizes to the DNA molecule.

[0278] At block 1920, one or more loci are identified. Each locus has a first allele on a first haplotype of a first chromosomal region. The first haplotype has a property of either: (1) not being present in healthy cells of the organism, but instead may be from a tumor or transplanted tissue, as examples; or (2) being present in normal cells of the organism and not being present in abnormal cells that may be in the mixture. Thus, the first haplotype has a genomic signature. In this manner, there is a difference between the healthy (normal) cells and the abnormal cells, thereby allowing embodiments to track a fractional contribution of one or the other, or both, so as to track an extent (e.g., fractional contribution) of the abnormal cells. With property (1), the first haplotype is associated with a disease state, e.g., cancer or a rejection of transplanted tissue. Thus, a particular cancer can have the first haplotype in a cancer genome of the particular cancer.

[0279] The one or more first alleles may be identified at the one or more loci on the first haplotype by obtaining a tissue sample (e.g., of the tumor or transplanted tissue) and analyzing DNA molecules of the tissue sample to determine the first haplotype. Such a tissue sample may be obtained from a biopsy, and method 1900 can be used to test if the cancer has metastasized to other tissues, or has recurred after surgery. Each one of the loci may be a heterozygous locus or a homozygous locus in the abnormal cells. For example, in FIG. 18, locus 1 and locus 3 are homozygous in the donor organ. But, ultimately, more than one allele would be observed in the plasma for all the loci, as each locus would have a signature for healthy cells or for abnormal cells. Thus, two haplotypes would exist across the tissue types, but a single tissue type might have only one haplotype in a region being analyzed.

[0280] At block 1930, a first set of the plurality of cell-free DNA molecules is identified. Each of the plurality of cell-free DNA molecules is located at any one of the loci from block 1920 and includes a corresponding first allele at the one locus, so that the cell-free DNA molecule can be identified as corresponding to the first haplotype. Each of the first set of cell-free DNA molecules also includes at least one of N genomic sites, where the genomic sites are used to measure the methylation levels. N is an integer, e.g., greater than or equal to 2, 3, 4, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, or 5,000.

[0281] At block 1940, N first mixture methylation levels are measured at the N genomic sites using the first set of the plurality of cell-free DNA molecules. One first mixture methylation level can be measured for each of the N genomic sites. Block 1940 can be performed in a similar

manner as block 150 of method 100 of FIG. 1. In some embodiments, the measurement of the methylation level of a DNA molecule can use methylation-aware sequencing results, which may also be used to determine the location and respective allele of the DNA molecule.

[0282] At block 1950, a first fractional contribution of the first tissue type in the mixture is determined using the N first methylation levels. In some embodiments, block 1950 can be performed via blocks 160 and 170 of method 100 of FIG. 1. Thus, a fractional contribution can be determined simultaneously for a panel of M tissue types. Block 1950 may use N issue-specific methylation levels at N genomic sites, determined for each of M tissue types, e.g., as in block 120 of method 100 of FIG. 1.

[0283] At block 1960, a separation value between the first fractional contribution and a reference fractional contribution is computed. Examples of separation values are described herein. The reference fractional contribution can be determined using samples from organisms that are healthy for the first tissue type. For a transplant example, the reference fractional contribution can be determined from one or more measurements of biological samples of organisms whose transplanted first tissue is not being rejected.

[0284] At block 1970, the separation value can be compared to a threshold value to determine a classification of whether the first tissue type has a disease state. For example, if the first haplotype is associated with cancer, then an appreciable first fractional contribution indicates that the first tissue type has cancer, as can be measured by the separation value exceeding the threshold value (e.g., when the reference fractional contribution is zero). The amount that the first fractional contribution exceeds the threshold value can indicate a certain level of cancer. As another example, the first haplotype can be specific to transplanted tissue, and a high contribution relative to the reference can indicate the organism is rejecting the transplanted tissue.

[0285] In an embodiment where the first haplotype is present in normal cells of the organism and not present in abnormal cells that may be in the mixture, the first tissue type can be determined to have the disease state when the first separation value is less than the threshold value. An example of a disease state is preeclampsia which can be associated with a spectrum of pathologic changes in a fetal tissue such as the placenta. As an example, in such a situation, if the first haplotype is specific to the fetus, e.g. paternally inherited haplotype, it may be increased in the maternal plasma in a pregnancy complicated with preeclampsia.

[0286] In some embodiments, a second haplotype for the diseased tissue, e.g., the transplanted tissue or a tumor may also be used. Thus, a second fractional contribution can be computed and compared to the reference fractional contribution. Accordingly, a second set of the plurality of cell-free DNA molecules can each be located at any one of the one or more loci, include a corresponding second allele on a second haplotype of the first chromosomal region, and include at least one of the N genomic sites. The second haplotype would have the same property of being only from healthy cells or abnormal cells.

[0287] A plurality of tissue types can be tested (e.g., using method 100 of FIG. 1), so as to determine a tissue of origin of the first haplotype, e.g., when it is associated with cancer. Accordingly, fractional contributions of other tissue types in the mixture can be determined using the N first methylation levels, and corresponding separation values between the corresponding fractional contributions and respective reference fractional contribution can be compared to the threshold value to determine a classification of whether each of the other tissue types has the particular cancer. Different tissue could have different reference fractional contributions.

VIII. IDENTIFYING TISSUE OF ORIGIN OF CNA OF CANCER

[0288] In some embodiments, an origin of a tumor may not be known. Thus, it can be difficult to identify point mutations in a tumor, as may be used for method 1900 of FIG. 19 or other methods described herein. Additionally, a tumor may not have a significant number of point mutations, but may have chromosomal regions exhibiting amplifications and deletions (examples of copy number aberrations).

[0289] To address this problem, embodiments can use a copy number analysis to identify regions that exhibit a copy number aberration (CNA). Typically, a CNA occurs on only one haplotype of a region. As only one haplotype has an amplification or a deletion, there will be a relatively large difference between the fractional contributions of the tissue type within which the tumor resides.

[0290] The CNA analysis may be performed in a variety of ways, e.g., as described in U.S. Patent Nos. 8,741,811 and 9,121,069. For example, the human genome (or genome for other type of organism) can be partitioned into approximately 3,000 non-overlapping 1-Mb bins. The number of reads mapping to each 1-Mb bin can be determined. After correcting for GC bias (Chen EZ, et

al. (2011) *PLoS One* 6(7):e21791), the sequence read density of each bin can be calculated. For each bin, the sequence read density of the test case can be compared to the values of the reference control subjects. Copy number gains and losses may be defined as 3 standard deviations above and below, respectively, the mean of the controls. Accordingly, identifying a first chromosomal region as exhibiting a copy number aberration can be based on a first amount of cell-free DNA molecules that are located in the first chromosomal region.

[0291] To determine the tissue origin of copy number aberrations in plasma, plasma DNA tissue mapping can be performed using the methylation markers located within the genomic regions exhibiting such aberrations in plasma. In the examples below for the cancer patients, mapping of plasma DNA copy number aberrations was performed only in cases with aberrations affecting a contiguous chromosome region of at least 30 Mb so that a sufficient number of methylation markers could be used for mapping.

A. Identifying regions with copy number aberration (CNA)

[0292] A 62-year-old male patient with HCC was recruited from the Department of Surgery, Prince of Wales Hospital, Hong Kong with informed consent. Ten milliliters of venous blood were collected in EDTA tubes at diagnosis and 3 months after the resection of the tumor. The blood samples were centrifuged at 3000 g for 10 minutes to separate the blood cells from the plasma. The plasma was recentrifuged at 30000 g for 10 minutes to remove the remaining cells.

[0293] DNA extracted from the blood cells was used for phasing the SNPs to construct the haplotypes of the patient using the 10x genomics platform following the manufacturer's instruction. High molecular weight DNA was extracted from blood or tissue samples using MagAttract HMW DNA kit (QIagen, Germany). The quality of DNA was verified by Genomic DNA Analysis ScreenTape on a 4200 TapeStation system (Agilent, Germany). DNA was quantified by dsDNA HS Assay kit on a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA). Sample indexing and library preparation were performed using the GemCode system and its associated reagents (10X Genomics, Pleasanton, CA) (Zheng et al. Nat Biotechnol. 2016 Mar;34:303-11). In brief, 1ng of DNA was inputted for GEM reactions in which individual DNA molecules were partitioned to introduce specific barcodes and extend the DNA. After GEM reactions, sequencing libraries were prepared according to the manufacturer's recommendations.

The libraries were quantified by qPCR using KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA). The normalized libraries were sequenced on a HiSeq 2500 sequencer (Illumina, San Diego, CA), with paired-end sequencing of 98-bp, 14-bp I5 and 8-bp I7 index reads. Sequencing results were analyzed using Long Ranger software suite (10X Genomics) so that all heterozygous SNPs were phased and the two haplotypes of the patient was determined.

[0294] The plasma samples were sequenced using the Illumina to a depth of 17x. Copy number aberrations were detected in the plasma of the HCC patient according to the method as previously described (Chan et al. Clin Chem. 2013;59:211-24).

[0295] FIG. 20 shows a plot of copy number aberrations detected in the plasma of an HCC patient according to embodiments of the present invention. The inner circle represents the result of the plasma sample collected at diagnosis (pre-operation) and the outer circle represents the result of the plasma sample collected at 3 months after the resection of the tumor (post-operation). Each dot represents a 1-Mb region. The green, red, and grey dots represent regions with copy number gain, copy number loss and no copy number change, respectively. Copy number aberrations were detected in the plasma sample at diagnosis and these changes disappeared after the tumor was removed.

[0296] In FIG. 20, two regions are highlighted for having a CNA. Region 2010 has a copy number gain, and region 2020 has a copy number loss. The haplotypes of these regions can be determined using any tissue sample of the subject, and not just a tumor sample. The difference in copy number is what is driving the difference in the fractional contributions, and that difference should be greatest in the tissue type with the tumor.

B. Determining the tissue origin of the copy number aberrations

[0297] We performed methylation deconvolution analysis for the two haplotypes independently. For illustration purposes, the two haplotypes are named Hap I and Hap II. The plasma DNA molecules covering heterozygous SNPs and at least one CpG site were used for this analysis. Plasma DNA molecules carrying the SNP alleles on Hap I were analyzed independently from those carrying alleles on Hap II. The methylation status of the CpG sites were used for methylation deconvolution for molecules mapped to Hap I and Hap II independently. As a result, the tissue contribution to Hap I and Hap II in plasma DNA could be determined.

[0298] First we focused on the regions with amplification. For illustration purposes, we analyzed the amplified region on chromosome 1q as an example.

	At diagnosis	After tumor resection
Hap I	34,119	11,131
Hap II	26,582	11,176

[0299] Table 11 shows the number of sequence reads from the two haplotypes. At diagnosis, the number of reads mapped to Hap I was increased compared with the number of reads mapped to Hap II. This indicated that Hap I is amplified relative to Hap II. This observation is compatible with the fact that a particular chromosome is duplicated in cancer rather than both homologous chromosomes being amplified to the same extent, which was in line with the fact that the copy number aberrations occurs preferentially on one haplotype (Adey A. et al, Nature. 2013;500:207-11;LaFramboise T. et al, PLoS Comput Biol. 2005;1(6):e65). The difference in the dosage of the two haplotypes disappeared after the tumor was resected. The difference in the absolute sequence reads number between the plasma samples taken at diagnosis and after tumor resection was due to the difference in the total number of sequence reads generated for the two plasma samples.

	At diagnosis			After tumor resection		
	Hap I	Hap II	Difference	Hap I	Hap II	Difference
liver	19.7	8.0	11.7	21.3	21.9	-0.6
Lung	5.4	0	5.4	0	0	0
Colon	0	0	0	0	0	0
Brain	0	0	0	9.0	9.0	0
Heart	0	17.0	-17	3.0	2.5	-0.5
Blood cell	74.9	75.0	0	66.7	66.6	0.1
Total	100	100	0	100	100	0

[0300] Table 12 shows a percentage contribution of different tissues to plasma DNA for the two haplotypes at diagnosis and after tumor resection. At diagnosis, the contributions of liver to plasma DNA were 19.7% and 8.0% for Hap I and Hap II, respectively. A difference of 11.7% was the highest among different types of tissue. This indicated that the dosage difference between Hap I and Hap II in the plasma was most likely contributed from the contribution of the liver. This further indicated that the likely origin of the chromosomal aberration was from the liver because the copy number changes was most likely due to the duplication of Hap I in the sequence read

counts analysis. In another embodiment, the difference in the contribution for Hap I and Hap II can be ranked to indicate the relative likelihood of different tissues being the source of the copy number aberrations.

[0301] The value for the heart is -17, which is in the opposite direction of the copy number aberration identified by Table 11. Thus, although the absolute value for the heart is larger than the absolute value for the liver, the opposite sign would discount the heart as being a viable candidate for the tissue type of the origin of the tumor. As the total contribution of all organs is 100%, the positive difference in the contributions of liver results in other tissues having negative values.

[0302] Similarly, this haplotype-specific methylation deconvolution can also be performed on regions with copy number loss. For illustration purpose, we performed this analysis on a region on chromosome 1p that exhibited copy number loss.

	At diagnosis	After tumor resection
Hap I	19,973	8,323
Hap II	12,383	7,724

[0303] Table 13 shows a number of sequence reads from the two haplotypes. At diagnosis, the number of reads mapped to Hap II was decreased compared with the number of reads mapped to Hap I. In tumor tissues, most of the regions with chromosome copy number loss would only involve the deletion of one of the two chromosomes. Thus the relative reduction in the dosage of Hap II was compatible with the deletion of Hap II. The difference in the dosage of the two haplotypes disappeared after the tumor was resected indicated that the amount of tumor-derived DNA had decreased or disappeared from the plasma.

	At diagnosis			After tumor resection		
	Hap I	Hap II	Difference (Hap I – Hap II)	Hap I	Hap II	Difference (Hap I – Hap II)
liver	13.3	5.5	7.8	10.2	13.2	-3
Lung	0	0	0	4.1	0.5	3.6
Colon	3.8	0	3.8	8.6	17.5	-8.9
Brain	0	0	0	0	0	0
Heart	3.7	0	3.7	25.5	19.4	6.1
Blood cell	79.2	94.5	-15.3	51.6	49.4	2.2
Total	100	100	0	100	100	0

[0304] Table 14 shows a percentage contribution of different tissues to plasma DNA for the two haplotypes at diagnosis and after tumor resection. At diagnosis, the contributions of liver to plasma DNA were 13.3% and 5.5% for Hap I and Hap II, respectively. A difference of 7.8% was the highest among different types of tissue. This indicated that the dosage difference between Hap I and Hap II in the plasma was most likely contributed from the contribution of the liver. This further indicated that the likely origin of the chromosomal aberration was from the liver because the copy number changes was most likely due to the deletion of Hap II in the sequence read counts analysis. In another embodiment, the difference in the contribution for Hap I and Hap II can be ranked to indicate the relative likelihood of different tissues being the source of the copy number aberrations.

C. Method of determining tissue origin of tumor

[0305] FIG. 21 is a flowchart illustrating a method of analyzing a biological sample of an organism to identify an origin of a chromosomal aberration according to embodiments of the present invention. The biological sample includes a mixture of cell-free DNA molecules from a plurality of tissues types that include a first tissue type.

[0306] At block 2110, a plurality of cell-free DNA molecules from the biological sample are analyzed. Block 2110 can be performed using techniques described in block 1910 of FIG. 1 and block 140 of method 100 of FIG. 1, as well as other blocks describing similar features.

[0307] At block 2115, a first chromosomal region is identified as exhibiting a copy number aberration in the organism based on a first amount of cell-free DNA molecules that are located in the first chromosomal region. As an example, plasma DNA analysis is performed to identify regions that exhibit copy number aberrations. The aberration can correspond to over or under-representation. In some embodiments, the genome can be separated into bins (e.g., 1-Mb bins), and the amount of cell-free DNA molecules from a particular bin can be determined (e.g., by mapping sequence reads to that part of a reference genome). The amount for a particular bin can be normalized (e.g., with respect to an average amount for a bin) can an over or under-representation can be identified.

[0308] Other techniques than counting DNA molecules mapping to a particular region can be used. For example, a distribution of sizes of the DNA molecules aligning to the first chromosomal

region can be used to detect the CNA. For example, cell-free tumor DNA is smaller than cell-free DNA from normal cells. This difference in size can be used to detect differences in a size distribution (e.g., average size or ratio of number of DNA molecules at different sizes) between two haplotypes for the region, or between the region and another region.

[0309] At block 2120, a first haplotype and a second haplotype of the organism in the first chromosomal region are determined. The two haplotypes may have been determined as part of block 2115. The two haplotypes can be determined using the same cell-free mixture or from a different sample, e.g., a cellular sample.

[0310] At block 2130, one or more heterozygous loci of the first chromosomal region are identified. Each heterozygous locus includes a corresponding first allele in the first haplotype and a corresponding second allele in the second haplotype. Block 2130 may be performed in a similar manner as other similar blocks of methods described herein.

[0311] At block 2140, a first set of the plurality of cell-free DNA molecules are identified. Each DNA molecule of the first set is located at any one of the one or more heterozygous loci, includes the corresponding first allele of the heterozygous locus, and includes at least one of N genomic sites. N is an integer greater than or equal to 2. Block 2140 may be performed in a similar manner as other similar blocks of methods described herein.

[0312] At block 2150, N first mixture methylation levels at the N genomic sites are measured using the first set of the plurality of cell-free DNA molecules. Block 2150 may be performed in a similar manner as other similar blocks of methods described herein.

[0313] At block 2160, a second set of the plurality of cell-free DNA molecules is identified. Each DNA molecule of the second set is located at any one of the one or more heterozygous loci, includes the corresponding second allele of the heterozygous locus, and includes at least one of the N genomic sites. Block 2160 may be performed in a similar manner as other similar blocks of methods described herein.

[0314] In some embodiments, a first number of cell-free DNA molecules in the first set of the plurality of cell-free DNA molecules can be determined, and a second number of cell-free DNA molecules in the second set of the plurality of cell-free DNA molecules can be determined, e.g., as shown in table 11. It can be determined which number is higher, thereby providing information

about the expected separation value for the tissue of origin, e.g., which haplotype should have a higher fractional contribution.

[0315] The first set of the plurality of cell-free DNA molecules can have a first size distribution, and the second set of the plurality of cell-free DNA molecules can have a second size distribution. A statistical value of a size distribution of the DNA molecules can be determined for each haplotype, thereby providing a first statistical value and a second statistical value. The haplotype with the smaller size distribution would be expected to have a higher copy number than the other haplotype, as tumor cell-free DNA is known to be smaller, as described in U.S. Patent No. 8,741,811. Examples of a statistical value of a size distribution are a ratio of a number of DNA molecules at different sizes, an average size, or a percentage of the DNA molecules at a specific size (e.g., below a size cutoff).

[0316] At block 2170, N second mixture methylation levels at the N genomic sites are measured using the second set of the plurality of cell-free DNA molecules. Block 2170 may be performed in a similar manner as other similar blocks of methods described herein.

[0317] Blocks 2180 and 2190 may be performed for each of a plurality of M tissue types. The M tissue types can include a default list of tissue types that are screened and for which reference methylation levels may be known. The default list can include tissues that cancer is most predominantly seen. M is an integer greater than one.

[0318] At block 2180, a computer system determines a corresponding first fractional contribution of the tissue type in the mixture using the N first methylation levels. The computer system determines a corresponding second fractional contribution of the tissue type in the mixture using the N second methylation levels. Block 2180 may be performed in a similar manner as other similar blocks of methods described herein.

[0319] At block 2190, a corresponding separation value between the corresponding first fractional contribution and the corresponding second fractional contribution is computed. Various separation values can be used, e.g., as are described herein.

[0320] At block 2195, the first tissue type is identified as being an origin of the copy number aberration based on a first separation value of the first tissue type having a maximal value among the corresponding separate values. The determination can require that the highest separation value

be sufficiently higher than the second highest separation value. For example, the difference can be required to be at least a threshold value, e.g., 1%, 2%, 35, 4%, 5%, 6%, or 7%. In one implementation, a difference between the first separation value and a next highest separation value can be compared to a threshold to determine a classification of how likely the first tissue type is the origin of the copy number aberration. Thus, even if the difference is not above the threshold, a probability or other classification can be provided. For instance, a linear relationship can be used from 0 to the threshold, where the probability is 100% once the difference is equal to the threshold.

[0321] Depending on how the separation value is determined, the maximal value can be a maximum negative number or a maximum positive number. For example, the difference values in table 14 could be determined using Hap II – Hap I. Whether the maximum should be a positive or a negative value can be determined using the analysis of DNA molecules on each haplotype, e.g., a count as in Table 13 or a size analysis as described above. In some implementations, the separation value can always be determined such that a maximum positive value is expected, e.g., by subtracting the fractional contribution of the haplotype with a lower copy number from the fractional contribution of the haplotype with the higher copy number.

[0322] After the origin is identified, an investigation using imaging modalities, e.g. computed tomography (CT) scan or magnetic resonance imaging (MRI), of the subject (entire subject or specifically of the candidate organ) can be performed to confirm or rule out the presence of a tumor in the organ. If presence of a tumor is confirmed, treatment can be performed, e.g., surgery (by a knife or by radiation) or chemotherapy.

IX. COMPUTER SYSTEM

[0323] Any of the computer systems mentioned herein may utilize any suitable number of subsystems. Examples of such subsystems are shown in FIG. 22 in computer apparatus 10. In some embodiments, a computer system includes a single computer apparatus, where the subsystems can be the components of the computer apparatus. In other embodiments, a computer system can include multiple computer apparatuses, each being a subsystem, with internal components. A computer system can include desktop and laptop computers, tablets, mobile phones and other mobile devices.

[0324] The subsystems shown in FIG. 22 are interconnected via a system bus 75. Additional subsystems such as a printer 74, keyboard 78, storage device(s) 79, monitor 76, which is coupled to display adapter 82, and others are shown. Peripherals and input/output (I/O) devices, which couple to I/O controller 71, can be connected to the computer system by any number of means known in the art such as input/output (I/O) port 77 (e.g., USB, FireWire®). For example, I/O port 77 or external interface 81 (e.g. Ethernet, Wi-Fi, etc.) can be used to connect computer system 10 to a wide area network such as the Internet, a mouse input device, or a scanner. The interconnection via system bus 75 allows the central processor 73 to communicate with each subsystem and to control the execution of a plurality of instructions from system memory 72 or the storage device(s) 79 (e.g., a fixed disk, such as a hard drive, or optical disk), as well as the exchange of information between subsystems. The system memory 72 and/or the storage device(s) 79 may embody a computer readable medium. Another subsystem is a data collection device 85, such as a camera, microphone, accelerometer, and the like. Any of the data mentioned herein can be output from one component to another component and can be output to the user.

[0325] A computer system can include a plurality of the same components or subsystems, e.g., connected together by external interface 81 or by an internal interface. In some embodiments, computer systems, subsystem, or apparatuses can communicate over a network. In such instances, one computer can be considered a client and another computer a server, where each can be part of a same computer system. A client and a server can each include multiple systems, subsystems, or components.

[0326] Aspects of embodiments can be implemented in the form of control logic using hardware (e.g. an application specific integrated circuit or field programmable gate array) and/or using computer software with a generally programmable processor in a modular or integrated manner. As used herein, a processor includes a single-core processor, multi-core processor on a same integrated chip, or multiple processing units on a single circuit board or networked. Based on the disclosure and teachings provided herein, a person of ordinary skill in the art will know and appreciate other ways and/or methods to implement embodiments of the present invention using hardware and a combination of hardware and software.

[0327] Any of the software components or functions described in this application may be implemented as software code to be executed by a processor using any suitable computer language

such as, for example, Java, C, C++, C#, Objective-C, Swift, or scripting language such as Perl or Python using, for example, conventional or object-oriented techniques. The software code may be stored as a series of instructions or commands on a computer readable medium for storage and/or transmission. A suitable non-transitory computer readable medium can include random access memory (RAM), a read only memory (ROM), a magnetic medium such as a hard-drive or a floppy disk, or an optical medium such as a compact disk (CD) or DVD (digital versatile disk), flash memory, and the like. The computer readable medium may be any combination of such storage or transmission devices.

[0328] Such programs may also be encoded and transmitted using carrier signals adapted for transmission via wired, optical, and/or wireless networks conforming to a variety of protocols, including the Internet. As such, a computer readable medium may be created using a data signal encoded with such programs. Computer readable media encoded with the program code may be packaged with a compatible device or provided separately from other devices (e.g., via Internet download). Any such computer readable medium may reside on or within a single computer product (e.g. a hard drive, a CD, or an entire computer system), and may be present on or within different computer products within a system or network. A computer system may include a monitor, printer, or other suitable display for providing any of the results mentioned herein to a user.

[0329] Any of the methods described herein may be totally or partially performed with a computer system including one or more processors, which can be configured to perform the steps. Thus, embodiments can be directed to computer systems configured to perform the steps of any of the methods described herein, potentially with different components performing a respective steps or a respective group of steps. Although presented as numbered steps, steps of methods herein can be performed at a same time or in a different order. Additionally, portions of these steps may be used with portions of other steps from other methods. Also, all or portions of a step may be optional. Additionally, any of the steps of any of the methods can be performed with modules, units, circuits, or other means for performing these steps.

[0330] The specific details of particular embodiments may be combined in any suitable manner without departing from the spirit and scope of embodiments of the invention. However, other

embodiments of the invention may be directed to specific embodiments relating to each individual aspect, or specific combinations of these individual aspects.

[0331] The above description of example embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form described, and many modifications and variations are possible in light of the teaching above.

[0332] A recitation of "a", "an" or "the" is intended to mean "one or more" unless specifically indicated to the contrary. The use of "or" is intended to mean an "inclusive or," and not an "exclusive or" unless specifically indicated to the contrary. Reference to a "first" component does not necessarily require that a second component be provided. Moreover reference to a "first" or a "second" component does not limit the referenced component to a particular location unless expressly stated.

[0333] All patents, patent applications, publications, and descriptions mentioned herein are incorporated by reference in their entirety for all purposes. None is admitted to be prior art.

WHAT IS CLAIMED IS:

1. A method of determining a portion of a fetal genome of an unborn fetus of a pregnant female using a biological sample from the pregnant female, wherein the biological sample including a mixture of cell-free DNA molecules from a plurality of tissues types, including maternal tissue types and a fetal tissue type, the fetus having a father and a mother being the pregnant female, the method comprising:

analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample, the plurality of cell-free DNA molecules being at least 1,000 cell-free DNA molecules, wherein analyzing a cell-free DNA molecule includes:

identifying a location of the cell-free DNA molecule in a reference human genome; and

determining a respective allele of the cell-free DNA molecule;

determining a first haplotype and a second haplotype of a first chromosomal region of a first parental genome of a first parent of the fetus;

identifying one or more heterozygous loci of the first chromosomal region of the first parental genome, each heterozygous locus including a corresponding first allele in the first haplotype and a corresponding second allele in the second haplotype;

identifying a first set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more heterozygous loci,

includes the corresponding first allele of the heterozygous locus, and

includes at least one of N genomic sites, N being an integer greater than or equal to 2;

measuring N first mixture methylation levels at the N genomic sites using the first set of the plurality of cell-free DNA molecules;

determining, by the computer system, a first fractional contribution of the fetal tissue type in the mixture using the N first methylation levels;

identifying a second set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more heterozygous loci,

includes the corresponding second allele, and

includes at least one of the N genomic sites;

measuring N second mixture methylation levels at the N genomic sites using the second set of the plurality of cell-free DNA molecules;

determining, by the computer system, a second fractional contribution of the fetal tissue type in the mixture using the N second methylation levels;

computing a first separation value between the first fractional contribution and the second fractional contribution;

determining the portion of fetal genome at the one or more heterozygous loci based on the first separation value.

2. The method of claim 1, wherein the one or more heterozygous loci are a first plurality of heterozygous loci.

3. The method of claim 2, wherein the first haplotype is a first maternal haplotype, and wherein the second haplotype is a second maternal haplotype.

4. The method of claim 3, further comprising:

identifying a second plurality of heterozygous loci of the first chromosomal region in a paternal genome, each of the second plurality of heterozygous loci including a corresponding third allele in a first paternal haplotype and a corresponding fourth allele in a second paternal haplotype, wherein the paternal genome corresponds to the father of the fetus;

identifying a third set of the plurality of cell-free DNA molecules that each:

is located at any one of the second plurality of heterozygous loci,

includes the corresponding third allele of the heterozygous locus, and

includes at least one of K genomic sites;

measuring K third mixture methylation levels at the K genomic sites using the third set of the second plurality of cell-free DNA molecules;

determining a third fractional contribution of the fetal tissue type in the mixture using the K third methylation levels;

identifying a fourth set of the plurality of cell-free DNA molecules that each:

is located at any one of the second plurality of heterozygous loci,

includes the corresponding fourth allele of the heterozygous locus, and

includes at least one of the K genomic sites;

measuring K fourth mixture methylation levels at the K genomic sites using the fourth set of the second plurality of cell-free DNA molecules;

determining a fourth fractional contribution of the fetal tissue type in the mixture using the K fourth methylation levels;

computing a second separation value between the third fractional contribution and the fourth fractional contribution;

determining the portion of fetal genome at the second plurality of heterozygous loci based on the second separation value.

5. The method of claim 2, wherein the first separation value is a ratio of the first fractional contribution and the second fractional contribution, and wherein the portion of the fetal genome is determined to have the first haplotype and the second haplotype when the ratio is equal to one within a cutoff value.

6. The method of claim 2, wherein the first separation value is a difference of the first fractional contribution and the second fractional contribution.

7. The method of claim 2, wherein the portion of the fetal genome is determined to have one or more copies of the first haplotype and no copies of the second haplotype when the first separation value is greater than a threshold value.

8. The method of claim 2, wherein the portion of the fetal genome is determined to have one or more copies of the second haplotype and no copies of the first haplotype when the first separation value is less than a threshold value.

9. The method of claim 1, wherein N is 10 or more.

10. The method of claim 1, wherein the N mixture methylation levels form a methylation vector b, and wherein determining the first fractional contribution of the fetal tissue type includes:

for each of M tissue types:

obtaining N tissue-specific methylation levels at the N genomic sites, N being greater than or equal to M, wherein the tissue-specific methylation levels form a matrix A of dimensions N by M, the M tissue types including the fetal tissue type;

solving for a composition vector x that provides the methylation vector b for the matrix A; and

for each of one or more components of the composition vector x:

using the component to determine a corresponding fractional contribution of a corresponding tissue type of the M tissue types in the mixture.

11. The method of claim 10, wherein the M tissue types include liver, lung, neutrophils, lymphocytes, erythroblasts, heart, large intestines, small intestines, and placenta, wherein the fetal tissue type is the placenta.

12. The method of claim 1, wherein measuring N first mixture methylation levels at the N genomic sites includes analyzing methylation-aware sequencing results, and wherein the locations of the plurality of cell-free DNA molecules are determined using the methylation-aware sequencing results.

13. The method of claim 1, wherein the first separation value includes a ratio of the first fractional contribution and the second fractional contribution.

14. The method of claim 1, wherein the first separation value includes a difference of the first fractional contribution and the second fractional contribution.

15. A method of determining a portion of a fetal genome of an unborn fetus of a pregnant female using a biological sample from the pregnant female, wherein the biological sample including a mixture of cell-free DNA molecules from a plurality of tissues types, including maternal tissue types and a fetal tissue type, the fetus having a father and a mother being the pregnant female, the method comprising:

analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample, the plurality of cell-free DNA molecules being at least 1,000 cell-free DNA molecules, wherein analyzing a cell-free DNA molecule includes:

identifying a location of the cell-free DNA molecule in a reference human genome; and

determining a respective allele of the cell-free DNA molecule;

determining a first haplotype of a first chromosomal region of a first parental genome of a first parent of the fetus;

identifying one or more heterozygous loci of the first chromosomal region of the first parental genome, each heterozygous locus including a corresponding first allele in the first haplotype and a corresponding second allele in a second haplotype of the first chromosomal region of the first parental genome;

identifying a first set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more heterozygous loci,

includes the corresponding first allele of the heterozygous locus, and

includes at least one of N genomic sites, N being an integer greater than or equal to 2;

measuring N first mixture methylation levels at the N genomic sites using the first set of the plurality of cell-free DNA molecules;

determining, by the computer system, a first fractional contribution of the fetal tissue type in the mixture using the N first methylation levels; and

comparing the first fractional contribution to a reference value to determine whether the fetus inherited the first haplotype at the first chromosomal region, wherein the fetus inherited the first haplotype at the first chromosomal region when the first fractional contribution exceeds the reference value.

16. The method of claim 15, wherein the reference value is determined by:

determining the second haplotype of the first chromosomal region of the first parental genome;

identifying a second set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more heterozygous loci,

includes the corresponding second allele of the heterozygous locus, and

includes at least one of the N genomic sites;

measuring N second mixture methylation levels at the N genomic sites using the second set of the plurality of cell-free DNA molecules;

measuring a second fractional contribution of the fetal tissue type in the mixture using the N second methylation levels;

determining the reference value as a sum of the second fractional contribution and a threshold value.

17. The method of claim 15, further comprising:

determining the second haplotype of the first chromosomal region of the first parental genome;

identifying a second set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more heterozygous loci,

includes the corresponding second allele of the heterozygous locus, and

includes at least one of the N genomic sites;

measuring N second mixture methylation levels at the N genomic sites using the second set of the plurality of cell-free DNA molecules;

determining a second fractional contribution of the fetal tissue type in the mixture using the N second methylation levels;

comparing the second fractional contribution to the reference value to determine whether the fetus inherited the second haplotype at the first chromosomal region, wherein the fetus inherited the second haplotype at the first chromosomal region when the second fractional contribution exceeds the reference value.

18. A method of determining a portion of a fetal genome of an unborn fetus of a pregnant female using a biological sample from the pregnant female, wherein the biological sample including a mixture of cell-free DNA molecules from a plurality of tissues types, including maternal tissue types and a fetal tissue type, the fetus having a father and a mother being the pregnant female, the method comprising:

analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample, the plurality of cell-free DNA molecules being at least 1,000 cell-free DNA molecules, wherein analyzing a cell-free DNA molecule includes:

identifying a location of the cell-free DNA molecule in a reference human genome; and

determining a respective allele of the cell-free DNA molecule;

determining a first haplotype and a second haplotype of a first chromosomal region of a first parental genome of a first parent of the fetus;

identifying a plurality of heterozygous loci of the first chromosomal region of the first parental genome, each heterozygous locus including a corresponding first allele in the first haplotype and a corresponding second allele in the second haplotype;

identifying a first set of the plurality of cell-free DNA molecules that each:

is located at any one of the plurality of heterozygous loci,

includes the corresponding first allele of the heterozygous locus, and

includes at least one of N genomic sites;

measuring a first mixture methylation level using the first set of the plurality of cell-free DNA molecules;

identifying a second set of the plurality of cell-free DNA molecules that each:

is located at any one of the plurality of heterozygous loci,

includes the corresponding second allele of the heterozygous locus, and

includes at least one of the N genomic sites;

measuring a second mixture methylation level using the second set of the plurality of cell-free DNA molecules; and

determining which of the first haplotype and the second haplotype is inherited by the fetus based on which of the first mixture methylation level and the second mixture methylation level is lower.

19. The method of claim 18, wherein the first mixture methylation level is a methylation density of the first set of the plurality of cell-free DNA molecules, and wherein the second mixture methylation level is the methylation density of the second set of the plurality of cell-free DNA molecules.

20. The method of claim 18, wherein determining which of the first haplotype and the second haplotype is inherited includes:

determining a separation value between the first mixture methylation level and the second mixture methylation level; and

comparing the separation value to a threshold value.

21 . The method of claim 20, wherein the threshold value is determined using Student's t-test, Mann-Whitney rank-sum test, or Chi-square test.

22. The method of claim 18, wherein determining the first haplotype and the second haplotype of the first chromosomal region of the first parental genome includes:

determining genotypes of the first parent genome at the plurality of heterozygous loci using a sample from the first parent;

obtaining a plurality of reference haplotypes; and

deducing the first haplotype and the second haplotype using the genotypes and the plurality of reference haplotypes.

23. A method of detecting a sequence imbalance in a portion of a fetal genome of an unborn fetus of a pregnant female using a biological sample from the pregnant female, wherein the biological sample including a mixture of cell-free DNA molecules from a plurality of tissues types, including maternal tissue types and a fetal tissue type, the fetus having a father and a mother being the pregnant female, the method comprising:

analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample, the plurality of cell-free DNA molecules being at least 1,000 cell-free DNA molecules, wherein analyzing a cell-free DNA molecule includes:

identifying a location of the cell-free DNA molecule in a reference human genome; and

determining a respective allele of the cell-free DNA molecule;

determining a first target haplotype of a target chromosomal region of a first parental genome of a first parent of the fetus;

identifying a plurality of target heterozygous loci of the target chromosomal region of the first parental genome, each target heterozygous locus including a corresponding first target allele in the first target haplotype and a corresponding second target allele in a second target haplotype of the target chromosomal region of the first parental genome;

identifying a target set of the plurality of cell-free DNA molecules that each:

is located at any one of the target heterozygous loci,

includes the corresponding first target allele of the target heterozygous locus, and

includes at least one of N genomic sites in the target chromosomal region, N being an integer greater than or equal to 2;

measuring N target mixture methylation levels at the N genomic sites using the target set of the plurality of cell-free DNA molecules;

determining, by the computer system, a first target fractional contribution of the fetal tissue type in the mixture using the N target methylation levels; and

determining a first reference haplotype of a reference chromosomal region of the first parental genome, the reference chromosomal region being different than the target chromosomal region;

identifying a plurality of reference heterozygous loci of the reference chromosomal region of the first parental genome, each reference heterozygous locus including a corresponding first reference allele in the first reference haplotype and a corresponding second reference allele in a second reference haplotype of the reference chromosomal region of the first parental genome;

identifying a reference set of the plurality of cell-free DNA molecules that each:

is located at any one of the reference heterozygous loci,

includes the corresponding first reference allele of the reference heterozygous locus, and

includes at least one of K genomic sites in the reference chromosomal region;

measuring K reference mixture methylation levels at the K genomic sites using the reference set of the plurality of cell-free DNA molecules;

determining, by the computer system, a first reference fractional contribution of the fetal tissue type in the mixture using the K reference methylation levels;

computing a first separation value between the first target fractional contribution and the first reference fractional contribution; and

comparing the first separation value to a threshold value to determine a classification of whether the fetus has the sequence imbalance for the target chromosomal region.

24. The method of claim 23, wherein the sequence imbalance is an amplification of the target chromosomal region.

25. The method of claim 23, wherein the threshold value used is dependent on a type of the sequence imbalance being tested.

26. The method of claim 23, further comprising:

measuring a fetal DNA percentage in the biological sample; and

using the fetal DNA percentage to determine the threshold value.

27. The method of claim 23, further comprising:

determining the threshold value using a reference group of samples, all of which have the sequence imbalance or do not have the sequence imbalance.

28. The method of claim 23, further comprising:

determining that the fetus has inherited the first target haplotype from the first parent; and

determining that the fetus has inherited the first reference haplotype from the first parent.

29. The method of claim 28, wherein determining that the fetus has inherited the first target haplotype from the first parent includes:

determining a second target fractional contribution of the fetal tissue type in the mixture corresponding to the second target haplotype;

computing a second separation value between the first target fractional contribution and the second target fractional contribution;

determining that the fetus has inherited the first target haplotype from the first parent based on the second separation value.

30. The method of claim 23, wherein the target chromosomal region is a first chromosome and the reference chromosomal region is a second chromosome different from the first chromosome.

31. The method of claim 23, wherein the first parent is the mother of the fetus.

32. A method of analyzing a biological sample of an organism, the biological sample including a mixture of cell-free DNA molecules from a plurality of tissues types, including a first tissue type, the method comprising:

analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample, the plurality of cell-free DNA molecules being at least 1,000 cell-free DNA molecules, wherein analyzing a cell-free DNA molecule includes:

identifying a location of the cell-free DNA molecule in a reference genome corresponding to the organism;

determining a respective allele of the cell-free DNA molecule;

identifying one or more loci, each having a corresponding first allele on a first haplotype of a first chromosomal region, wherein the first haplotype is:

not present in healthy cells of the organism, or

present in the healthy cells of the organism and not present in abnormal cells that may be in the mixture;

identifying a first set of the plurality of cell-free DNA molecules that each:

is located at any one locus of the one or more loci,

includes the corresponding first allele of the one locus, and

includes at least one of N genomic sites, N being an integer greater than or equal to 2;

measuring N first mixture methylation levels at the N genomic sites using the first set of the plurality of cell-free DNA molecules;

determining, by the computer system, a first fractional contribution of the first tissue type in the mixture using the N first methylation levels;

determining a first separation value between the first fractional contribution and a reference fractional contribution; and

comparing the first separation value to a threshold value to determine a classification of whether the first tissue type has a disease state.

33. The method of claim 32, wherein the first haplotype is on a first chromosome, and wherein the first tissue type has the first haplotype on both copies of the first chromosome.

34. The method of claim 32, wherein the first haplotype is not present in the healthy cells of the organism, and wherein the first haplotype is associated with the disease state.

35. The method of claim 34, wherein the first tissue type is determined to have the disease state when the first separation value is greater than the threshold value.

36. The method of claim 34, further comprising:

identifying a second set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more loci;

includes a corresponding second allele on a second haplotype of the first chromosomal region, the second haplotype not being present in the healthy cells of the organism, and the second haplotype associated with the disease state; and

includes at least one of the N genomic sites;

measuring N second mixture methylation levels at the N genomic sites using the second set of the plurality of cell-free DNA molecules;

determining, by the computer system, a second fractional contribution of the first tissue type in the mixture using the N second methylation levels;

determining a second separation value between the second fractional contribution and the reference fractional contribution; and

comparing the second separation value to the threshold value as part of determining the classification of whether the first tissue type has the disease state.

37. The method of claim 34, wherein the disease state is cancer.

38. The method of claim 37, wherein a particular cancer has the first haplotype in a cancer genome of the particular cancer.

39. The method of claim 38, wherein the particular cancer is selected from a group consisting of liver cancer, lung cancer, pancreas cancer, atrium cancer, colon cancer, sigmoid colon cancer, transverse colon cancer, ascending colon cancer, descending colon cancer, adrenal gland cancer, esophagus cancer, small intestines cancer, and CD4 T cell cancer.

40. The method of claim 37, wherein the classification of whether the first tissue type has cancer comprises classification of the presence or absence of a cancer, classification of a stage of cancer, classification of a size of a tumor, and/or classification of metastasis.

41. The method of claim 37, further comprising performing, by the computer system:

determining a plurality of corresponding fractional contributions of other tissue types in the mixture using the N first mixture methylation levels;

determining corresponding separation values between the plurality of corresponding fractional contributions and corresponding reference fractional contributions; and

comparing the corresponding separation values to the threshold value to determine classifications of whether each of the other tissue types has the particular cancer.

42. The method of claim 34, wherein the corresponding first allele at each of the one or more loci is a cancer-specific mutation.

43. The method of claim 32, further comprising:

determining the corresponding first allele at each of the one or more loci on the first haplotype by:

analyzing DNA molecules of a tissue sample having the first haplotype to determine the first haplotype.

44. The method of claim 32, wherein the disease state is a rejection of a transplant of the first tissue type in the organism.

45. The method of claim 44, wherein the reference fractional contribution is determined from one or more measurements of biological samples of organisms with the first tissue type being healthy.

46. The method of claim 44, wherein the reference fractional contribution is determined from one or more measurements of biological samples of organisms whose transplanted first tissue is not being rejected.

47. The method of claim 32, wherein the reference fractional contribution is zero.

48. The method of claim 32, wherein the N first mixture methylation levels form a methylation vector b, and wherein determining the first fractional contribution of the first tissue type includes:

for each of M tissue types:

obtaining N tissue-specific methylation levels at the N genomic sites, N being greater than or equal to M, wherein the tissue-specific methylation levels form a matrix A of dimensions N by M, the M tissue types including the first tissue type, M being an integer greater than one;

solving for a composition vector x that provides the methylation vector b for the matrix A

for each component of one or more components of the composition vector x:

using the component to determine a corresponding fractional contribution of a corresponding tissue type of the M tissue types in the mixture.

49. The method of claim 32, wherein the first haplotype is present in the healthy cells of the organism and not present in the abnormal cells that may be in the mixture.

50. The method of claim 32, wherein the first tissue type is determined to have the disease state when the first separation value is less than the threshold value.

51. The method of claim 49, wherein the abnormal cells are from a tumor.

52. The method of claim 49, wherein the abnormal cells are from donor tissue.

53. A method of analyzing a biological sample of an organism to identify an origin of a chromosomal aberration, the biological sample including a mixture of cell-free DNA molecules from a plurality of tissues types, including a first tissue type, the method comprising:

analyzing, by a computer system, a plurality of cell-free DNA molecules from the biological sample, the plurality of cell-free DNA molecules being at least 1,000 cell-free DNA molecules, wherein analyzing a cell-free DNA molecule includes:

identifying a location of the cell-free DNA molecule in a reference genome corresponding to the organism;

determining a respective allele of the cell-free DNA molecule;

identifying a first chromosomal region as exhibiting a copy number aberration in the organism based on a first amount of cell-free DNA molecules that are located in the first chromosomal region;

determining a first haplotype and a second haplotype of the organism in the first chromosomal region;

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

identifying a first set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more heterozygous loci,

includes the corresponding first allele of the heterozygous locus, and

includes at least one of N genomic sites, N being an integer greater than or equal to 2;

measuring N first mixture methylation levels at the N genomic sites using the first set of the plurality of cell-free DNA molecules;

identifying a second set of the plurality of cell-free DNA molecules that each:

is located at any one of the one or more heterozygous loci,

includes the corresponding second allele of the heterozygous locus, and

includes at least one of the N genomic sites;

measuring N second mixture methylation levels at the N genomic sites using the second set of the plurality of cell-free DNA molecules;

for each tissue type of M tissue types:

determining, by the computer system, a corresponding first fractional contribution of the tissue type in the mixture using the N first methylation levels, M being an integer greater than one;

determining, by the computer system, a corresponding second fractional contribution of the tissue type in the mixture using the N second methylation levels;

computing a corresponding separation value between the corresponding first fractional contribution and the corresponding second fractional contribution;

identifying the first tissue type as being the origin of the copy number aberration based on a first separation value of the first tissue type having a maximal value among the corresponding separate values.

54. The method of claim 53, further comprising:

determining a difference between the first separation value and a next highest separation value; and

comparing the difference to a threshold to determine a classification of how likely the first tissue type is the origin of the copy number aberration.

55. The method of claim 53, wherein identifying the first chromosomal region as exhibiting the copy number aberration includes:

comparing the first amount of cell-free DNA molecules that are located in the first chromosomal region to a reference value determined from amounts of cell-free DNA molecules that are located in other chromosomal regions.

56. The method of claim 53, wherein the copy number aberration is a copy number gain.

57. The method of claim 53, further comprising:

determining a first number of cell-free DNA molecules in the first set of the plurality of cell-free DNA molecules;

determining a second number of cell-free DNA molecules in the second set of the plurality of cell-free DNA molecules; and

using the first number and the second number to determine whether the maximal value is to be positive or negative.

58. The method of claim 53, wherein analyzing the cell-free DNA molecule includes: determining a size of the cell-free DNA molecule, wherein the first set of the plurality of cell-free DNA molecules has a first size distribution, and wherein the second set of the plurality of cell-free DNA molecules has a second size distribution, the method further comprising:

determining a first statistical value of the first size distribution;

determining a second statistical value of the second size distribution; and

using the first statistical value and the second statistical value to determine whether the maximal value is to be positive or negative.

59. The method of claim 58, further comprising:

determining which of the first haplotype and the second haplotype has a higher copy number using the first statistical value and the second statistical value; and

computing the corresponding separation values such that the first separation value is positive.

60. The method of any one of claims 1-59, wherein analyzing the cell-free DNA molecules comprises bisulfite sequencing, sequencing preceded by methylation-sensitive restriction enzyme digestion, immunoprecipitation using anti-methylcytosine antibody or methylation binding protein, or single molecule sequencing that allows elucidation of the methylation status.

61. A computer product comprising a computer readable medium storing a plurality of instructions for controlling a computer system to perform an operation of any method of claims 1-59.

62. A system comprising:

the computer product of claim 61; and

one or more processors for executing instructions stored on the computer readable medium.

63. A system comprising means for performing any method of claims 1-59 .

64. A system configured to perform any method of claims 1-59.

65. A system comprising modules that respectively perform the steps of any method of claims 1-59.

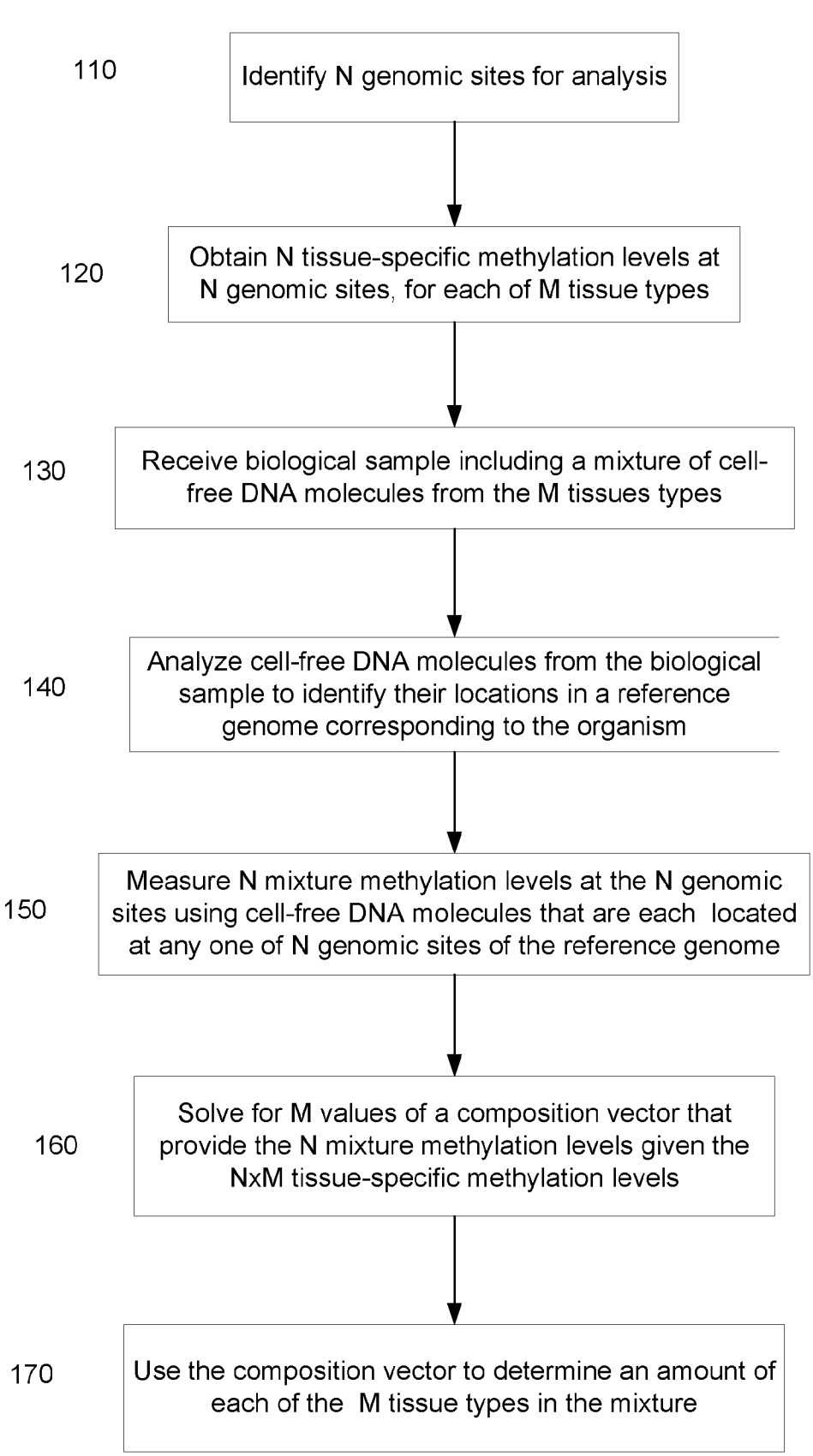

100
110
Identify N genomic sites for analysis
120
Obtain N tissue-specific methylation levels at N genomic sites, for each of M tissue types
130
Receive biological sample including a mixture of cell-free DNA molecules from the M tissues types
140
Analyze cell-free DNA molecules from the biological sample to identify their locations in a reference genome corresponding to the organism
150
Measure N mixture methylation levels at the N genomic sites using cell-free DNA molecules that are each located at any one of N genomic sites of the reference genome
160
Solve for M values of a composition vector that provide the N mixture methylation levels given the NxM tissue-specific methylation levels
170
Use the composition vector to determine an amount of each of the M tissue types in the mixture

FIG. 1

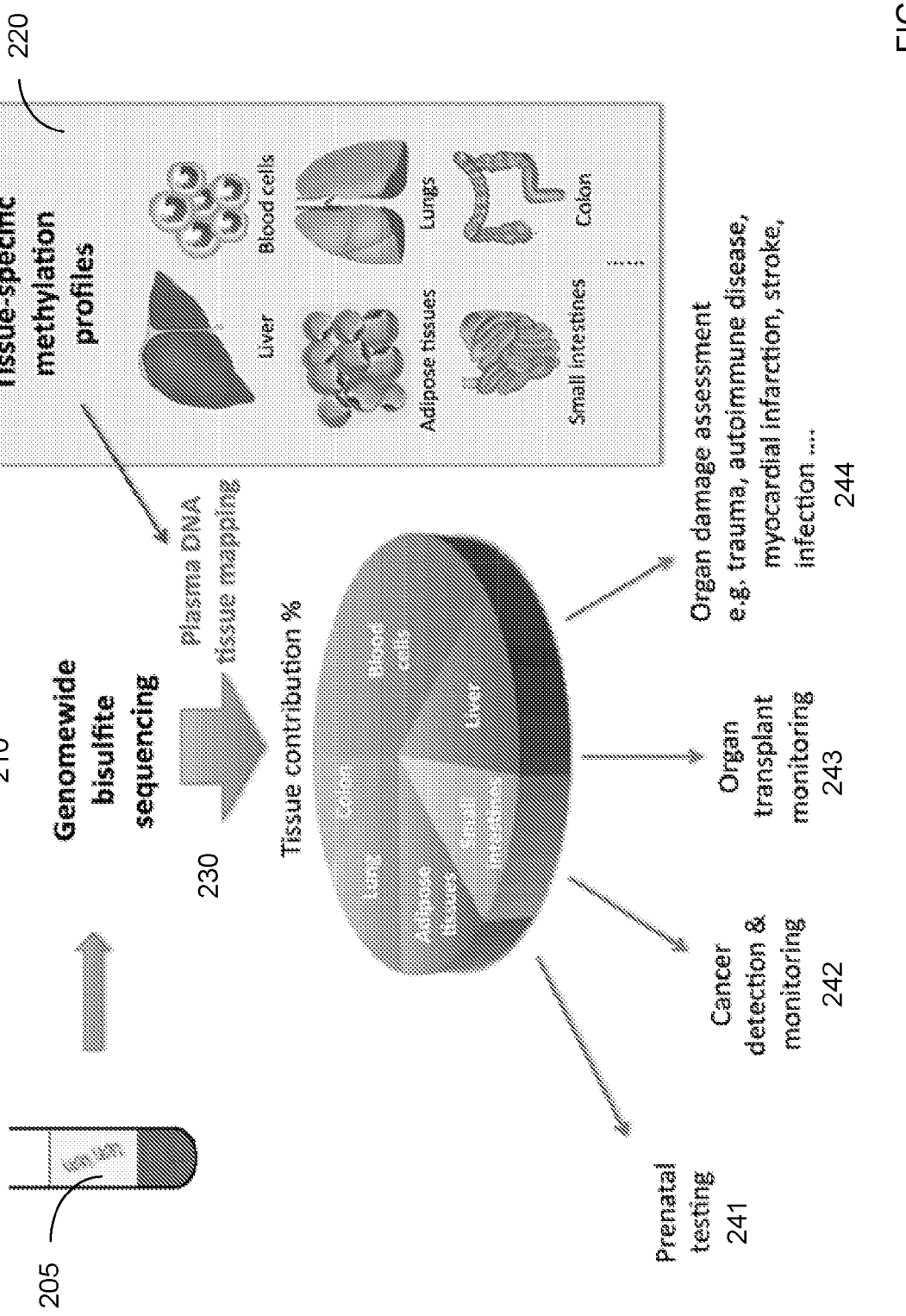

205
210
Genomewide bisulfite sequencing
230
Plasma DNA tissue mapping
220
Tissue-specific methylation profiles
Liver
Blood cells
Adipose tissues
Lungs
Small intestines
Colon
Tissue contribution %
Prenatal testing 241
Cancer detection & monitoring 242
Organ transplant monitoring 243
Organ damage assessment e.g. trauma, autoimmune disease, myocardial infarction, stroke, infection 244

FIG. 2

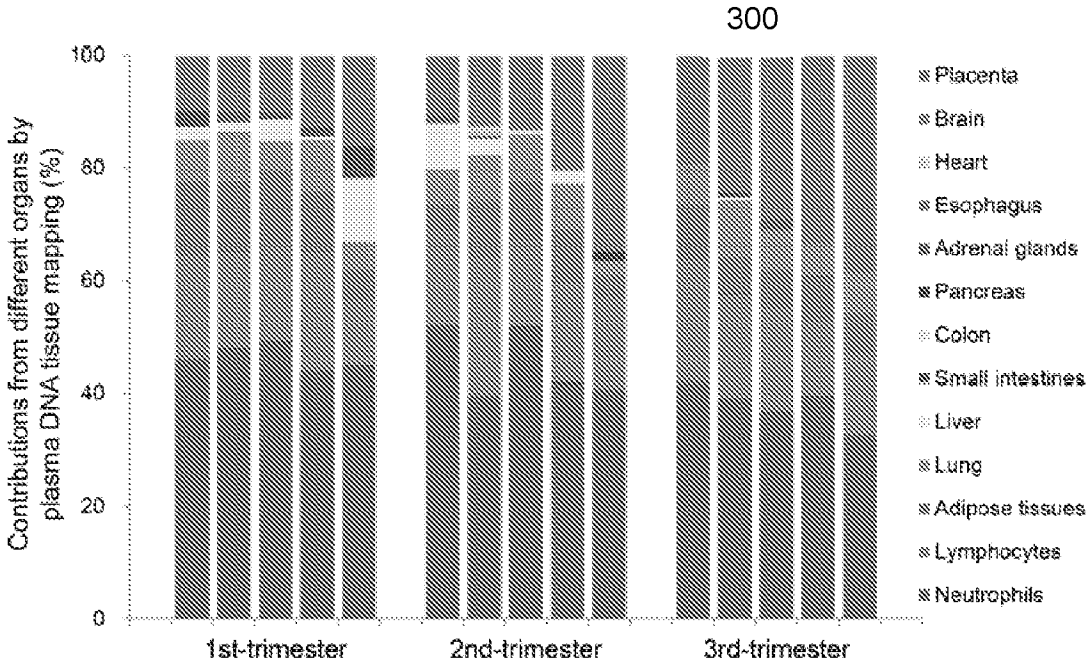

300
Contributions from different organs by plasma DNA tissue mapping (%)
100
80
60
40
20
0
1st-trimester
2nd-trimester
3rd-trimester
Placenta
Brain
Heart
Esophagus
Adrenal glands
Pancreas
Colon
Small intestines
Liver
Lung
Adipose tissues
Lymphocytes
Neutrophils

FIG. 3A

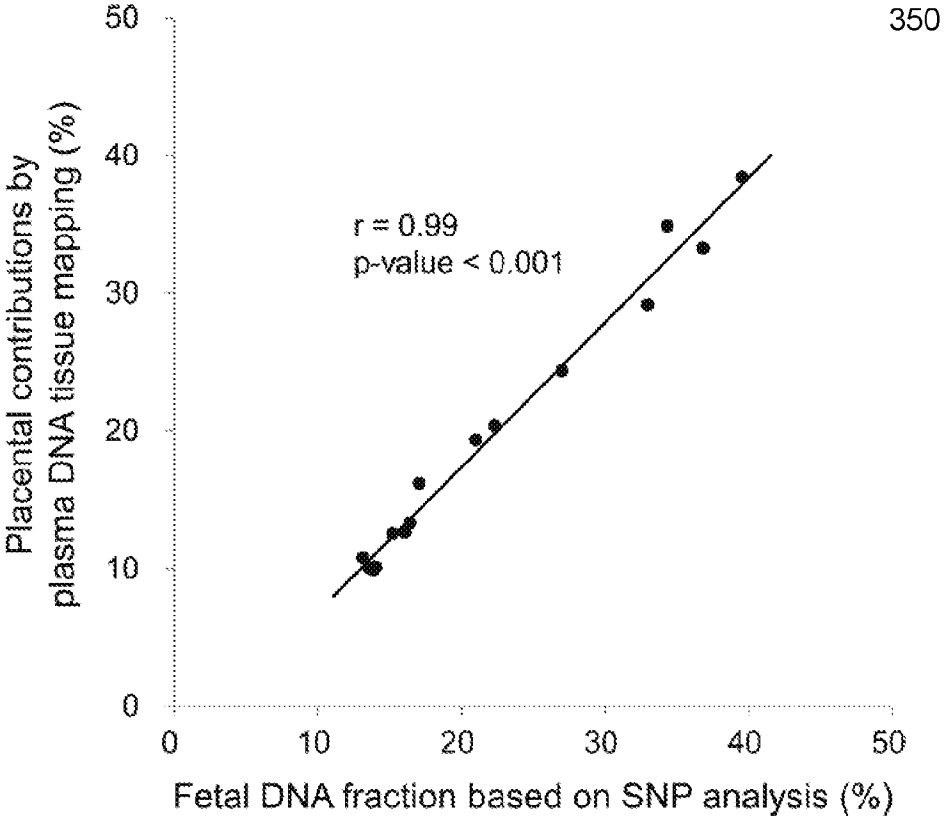

350
Placental contributions by plasma DNA tissue mapping (%)
50
40
30
20
10
0
r = 0.99
p-value < 0.001
0
10
20
30
40
50
Fetal DNA fraction based on SNP analysis (%)

FIG. 3B

Case no.	Percentage contributions of plasma DNA (%)													SNP-based fetal DNA fraction (%)	Gestational age (weeks)
	Liver	Lungs	Colon	Small intestines	Pancreas	Adrenal glands	Esophagus	Adipose tissues	Heart	Brain	Neutrophils	Lymphocytes	Placenta		
1	1.5	9.3	0.0	0.0	0.0	1.9	0.0	12.1	0.0	0.0	48.1	17.0	10.0	13.6	13 1/7
2	2.5	7.6	0.0	2.8	0.0	0.0	0.0	12.6	0.0	0.0	46.1	18.6	9.9	13.9	13
3	3.8	5.8	0.0	0.7	0.0	0.6	0.0	12.4	0.0	0.0	49.2	17.4	10.1	14.0	12 6/7
4	0.6	9.2	0.0	1.8	0.0	0.0	0.0	10.6	0.0	0.0	44.1	21.1	12.7	16.1	13
5	11.1	5.2	0.0	5.7	0.0	0.0	0.0	5.1	0.0	0.0	45.0	11.7	16.2	17.1	12 6/7
6	8.1	6.4	0.0	0.0	0.0	1.3	0.0	4.5	0.0	0.0	51.9	17.0	10.8	13.1	21 3/7
7	3.0	8.2	0.0	0.0	0.0	0.5	0.0	10.1	1.7	0.0	39.5	24.5	12.5	15.2	20 6/7
8	0.8	8.5	0.0	0.0	0.0	0.0	0.0	10.2	0.0	0.0	51.7	15.5	13.3	16.4	22 2/7
9	2.6	8.0	0.0	0.0	0.0	0.0	0.0	9.2	0.0	0.0	41.7	18.1	20.4	22.3	22 1/7
10	0.0	2.6	0.0	0.0	1.6	0.0	0.0	4.1	0.0	0.0	40.5	16.3	34.9	34.3	21 /7
11	0.0	7.4	0.0	0.0	0.0	0.0	0.0	9.8	0.0	0.0	41.5	22.0	19.3	21.0	38 1/7
12	0.6	10.1	0.0	0.8	0.0	0.0	0.0	6.6	0.0	0.0	39.0	18.7	24.3	27.0	38 3/7
13	0.0	7.4	0.0	0.0	0.0	1.7	0.0	6.3	0.0	0.0	36.7	18.7	29.2	33.0	38 1/7
14	0.0	5.9	0.0	0.0	0.0	0.0	0.0	8.3	0.0	0.0	39.6	12.9	33.3	36.8	38 2/7
15	0.0	7.3	0.0	0.0	0.0	0.0	0.0	7.2	0.0	0.0	32.5	14.6	38.4	39.5	38 3/7

FIG. 4

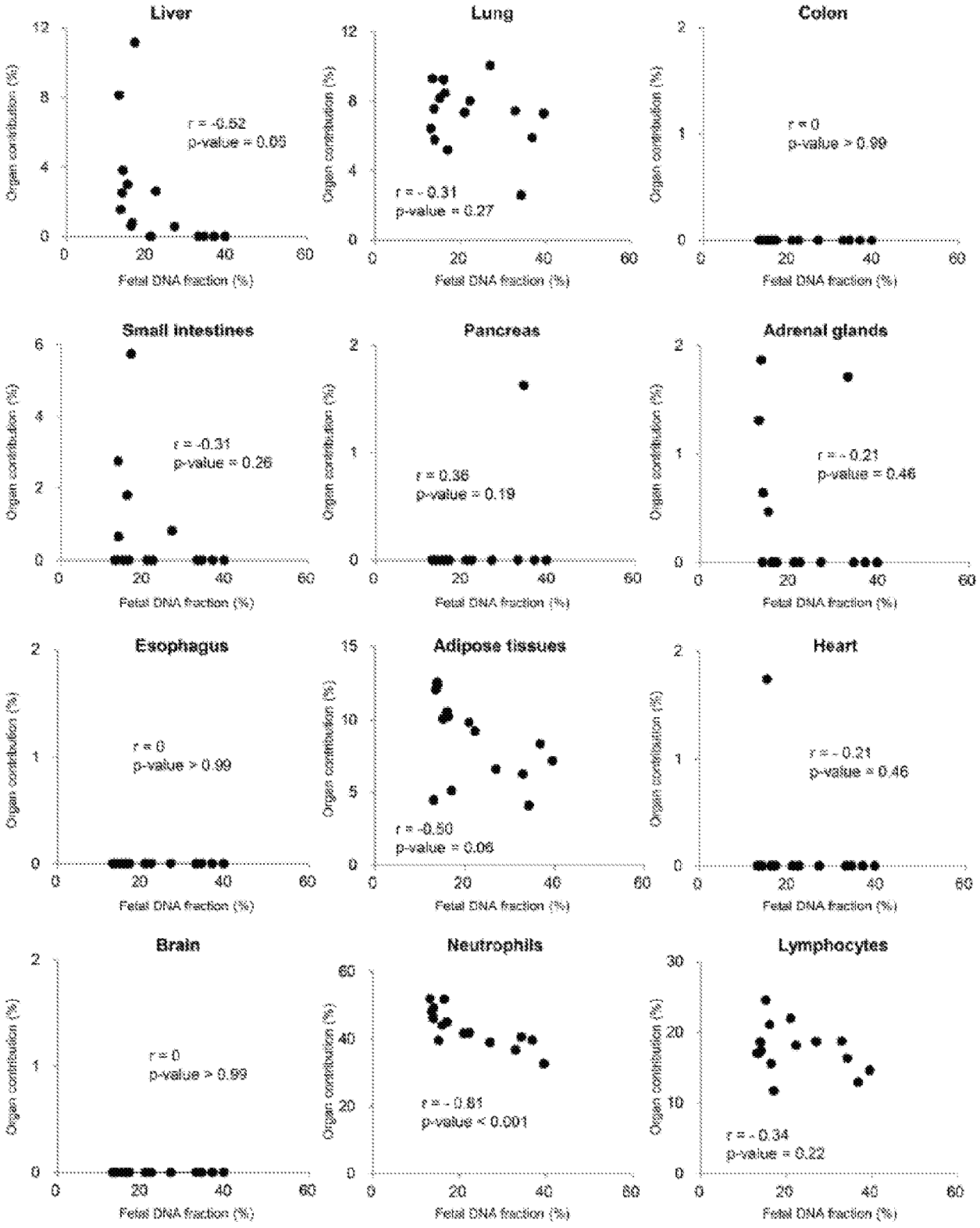

Liver
r = -0.52
p-value = 0.05
Lung
r = - 0.31
p-value = 0.27
Colon
r = 0
p-value > 0.99
Small intestines
r = -0.31
p-value = 0.26
Pancreas
r = 0.38
p-value = 0.19
Adrenal glands
r = - 0.21
p-value = 0.46
Esophagus
r = 0
p-value > 0.99
Adipose tissues
r = -0.50
p-value = 0.06
Heart
r = - 0.21
p-value = 0.46
Brain
r = 0
p-value > 0.99
Neutrophils
r = - 0.91
p-value < 0.001
Lymphocytes
r = - 0.34
p-value = 0.22
Organ contribution (%)
Fetal DNA fraction (%)

FIG. 5

600

Case no.	Percentage contributions of plasma DNA (%)												
	Liver	Lungs	Colon	Small intestines	Pancreas	Adrenal glands	Esophagus	Adipose tissues	Heart	Brain	Neutrophils	Lymphocytes	Placenta
1	5.8	6.8	0.0	1.3	0.0	0.0	0.0	6.2	2.9	0.0	58.2	18.8	0.0
2	5.1	5.2	0.0	7.3	0.0	0.0	0.0	4.5	0.0	0.0	61.8	16.1	0.0
3	5.6	7.8	0.0	2.1	0.0	0.0	0.0	6.5	0.6	0.0	58.8	18.6	0.0
4	4.1	6.4	0.0	4.5	0.0	0.0	0.0	5.9	1.3	0.4	59.0	18.4	0.0
5	4.4	2.5	0.7	6.4	0.0	0.0	0.0	5.5	2.7	0.6	59.6	17.1	0.3
6	5.8	4.1	0.0	5.3	0.0	0.0	0.0	5.2	0.5	0.8	55.2	22.7	0.5
7	4.9	4.1	0.0	4.5	0.0	0.0	0.0	6.3	0.0	0.0	60.2	20.0	0.0
8	4.4	7.4	0.0	5.3	0.0	0.0	0.0	3.1	0.3	0.3	61.7	17.5	0.0
9	4.5	9.4	0.0	4.0	0.0	0.0	0.0	6.2	0.1	1.2	55.9	18.6	0.0
10	4.6	0.1	0.2	1.3	0.0	0.0	0.0	7.5	0.0	2.4	60.0	23.9	0.0
11	4.7	4.5	0.0	5.5	0.0	0.0	0.0	5.6	0.9	0.5	62.4	16.0	0.0
12	7.3	4.9	0.0	4.4	0.0	0.0	0.0	6.0	1.8	0.9	54.0	20.7	0.0
13	2.4	10.9	0.0	1.7	0.0	0.0	0.0	5.7	0.0	0.6	53.9	24.5	0.3
14	3.8	3.5	0.0	0.0	0.0	0.0	0.0	8.9	0.0	0.0	61.2	22.7	0.0
15	5.3	7.2	0.0	4.2	0.0	0.2	0.0	5.1	0.0	0.1	56.3	21.6	0.0
16	4.9	6.5	0.0	7.9	0.0	0.0	0.0	4.3	0.5	1.2	55.4	18.9	0.4
17	18.6	12.8	0.0	0.0	0.0	0.0	0.0	4.4	0.0	0.0	47.7	16.5	0.0
18	8.3	5.0	0.0	2.1	0.0	0.0	0.0	7.5	0.0	0.2	56.2	19.4	1.3
19	6.0	0.0	1.0	2.3	0.0	0.0	0.0	9.3	0.0	0.0	54.5	26.3	0.6
20	2.0	4.8	0.0	0.0	0.0	0.5	0.0	10.5	0.0	0.0	58.5	22.5	1.1
21	8.5	0.0	0.0	4.5	0.0	0.0	0.0	9.4	0.0	0.0	53.9	23.4	0.2
22	6.4	11.1	0.0	4.0	0.0	0.0	0.0	6.7	0.0	0.0	53.7	18.1	0.0
23	2.0	2.8	0.0	8.7	0.0	0.0	0.0	6.5	1.1	0.0	62.5	15.0	1.3
24	5.4	2.2	0.0	2.7	0.0	0.0	0.0	6.9	0.0	1.6	65.2	15.2	0.9
25	21.0	0.0	0.0	7.1	0.0	0.0	0.0	3.8	0.0	0.0	51.6	16.4	0.0
26	7.0	6.4	0.0	6.1	0.0	0.0	0.0	4.1	1.6	0.0	55.9	19.0	0.0
27	7.9	10.4	0.0	0.3	0.0	0.0	0.0	4.8	0.0	0.0	59.1	17.6	0.0
28	5.4	9.7	0.0	4.2	0.0	0.0	0.0	5.1	0.0	0.0	55.7	20.0	0.0
29	6.2	6.6	0.0	0.0	0.0	0.0	0.0	6.3	0.0	0.0	60.7	20.2	0.0
30	6.9	5.1	0.0	5.2	0.0	0.0	0.0	6.4	1.4	0.0	58.8	16.2	0.0
31	14.0	6.4	0.0	4.5	0.0	0.0	0.0	4.9	0.7	0.0	53.2	16.2	0.0
32	9.9	5.5	0.0	1.9	0.0	0.0	0.0	6.3	0.6	0.0	60.1	15.6	0.0
Median	5.5	5.3	0.0	4.2	0.0	0.0	0.0	6.1	0.0	0.0	58.3	18.7	0.0
Lower quantile	4.6	3.9	0.0	1.9	0.0	0.0	0.0	5.0	0.0	0.0	55.0	16.5	0.0
Upper quantile	7.1	7.3	0.0	5.3	0.0	0.0	0.0	6.6	0.8	0.5	60.1	21.0	0.3

FIG. 6

700

	Pregnant women											Non-pregnant subjects			
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4
Buffy coat	58%	64%	65%	53%	83%	52%	75%	71%	52%	66%	69%	87%	83%	96%	80%
Esophagus	0%	0%	0%	2%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%
Small intestines	0%	0%	0%	4%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Colon	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	5%	0%	0%
Pancreas	0%	1%	0%	1%	0%	0%	0%	0%	1%	3%	2%	1%	0%	0%	0%
Liver	3%	8%	10%	1%	0%	11%	6%	7%	4%	9%	8%	7%	7%	3%	6%
Lung	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Heart	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	2%
Adrenal gland	0%	1%	0%	1%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	5%
Hippocampus	1%	3%	0%	1%	2%	4%	1%	4%	3%	3%	4%	2%	5%	0%	5%
Placenta	38%	24%	25%	36%	15%	34%	18%	15%	40%	19%	7%	2%	0%	1%	2%
Fetal DNA fraction determined by counting fetal specific alleles	32%	21%	26%	28%	11%	25%	17%	13%	13%	32%	3%	0%	0%	0%	0%

FIG. 7

800

	Pregnant women											Non-pregnant subjects			
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4
Buffy coat	43%	44%	47%	40%	47%	41%	38%	41%	43%	39%	45%	10%	11%	1%	13%
Esophagus	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Small intestines	0%	4%	1%	2%	3%	0%	1%	5%	0%	0%	0%	0%	0%	0%	0%
Colon	25%	18%	10%	23%	14%	23%	24%	20%	27%	25%	16%	17%	7%	22%	27%
Pancreas	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Liver	4%	8%	11%	4%	8%	11%	9%	0%	0%	0%	0%	0%	0%	0%	0%
Lung	0%	0%	0%	1%	0%	0%	0%	3%	0%	0%	11%	0%	0%	0%	0%
Heart	0%	0%	1%	3%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
Adrenal gland	0%	3%	6%	0%	0%	0%	7%	4%	3%	7%	7%	24%	23%	20%	4%
Hippocampus	7%	10%	7%	11%	12%	10%	8%	15%	9%	11%	16%	44%	42%	41%	50%
Placenta	21%	13%	18%	16%	15%	16%	11%	12%	18%	18%	6%	4%	16%	15%	7%
Fetal DNA fraction determined by counting fetal specific alleles	32%	21%	26%	28%	11%	25%	17%	13%	13%	32%	3%	0%	0%	0%	0%

FIG. 8

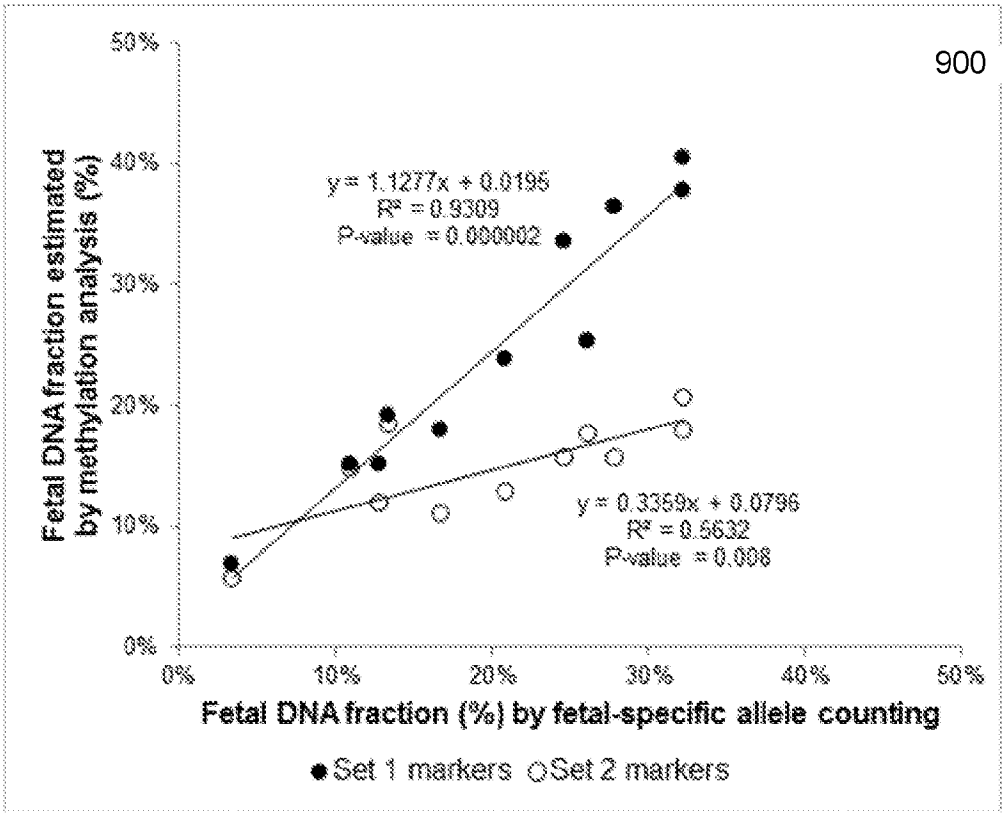

900
50%
40%
30%
20%
10%
0%
Fetal DNA fraction estimated by methylation analysis (%)
y = 1.1277x + 0.0195
R² = 0.9309
P-value = 0.000002
y = 0.3359x + 0.0796
R² = 0.5632
P-value = 0.008
0%
10%
20%
30%
40%
50%
Fetal DNA fraction (%) by fetal-specific allele counting
Set 1 markers
Set 2 markers

FIG. 9A

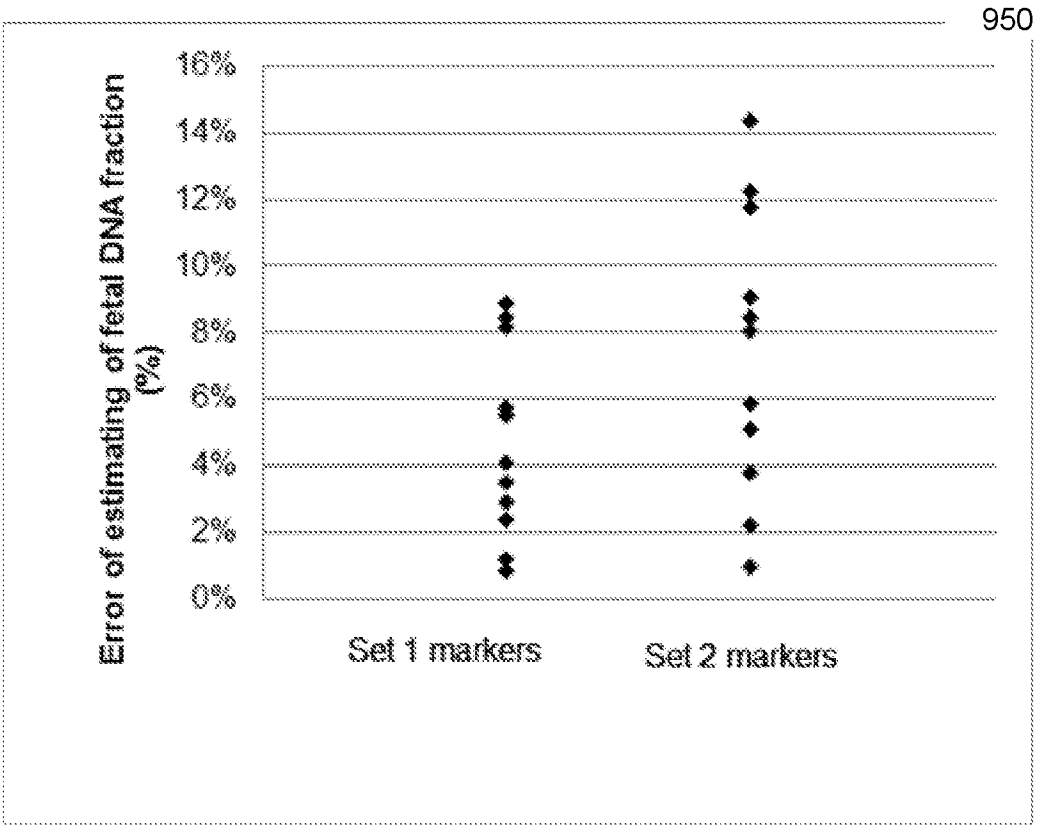

950
16%
14%
12%
10%
8%
6%
4%
2%
0%
Error of estimating of fetal DNA fraction (%)
Set 1 markers
Set 2 markers

FIG. 9B

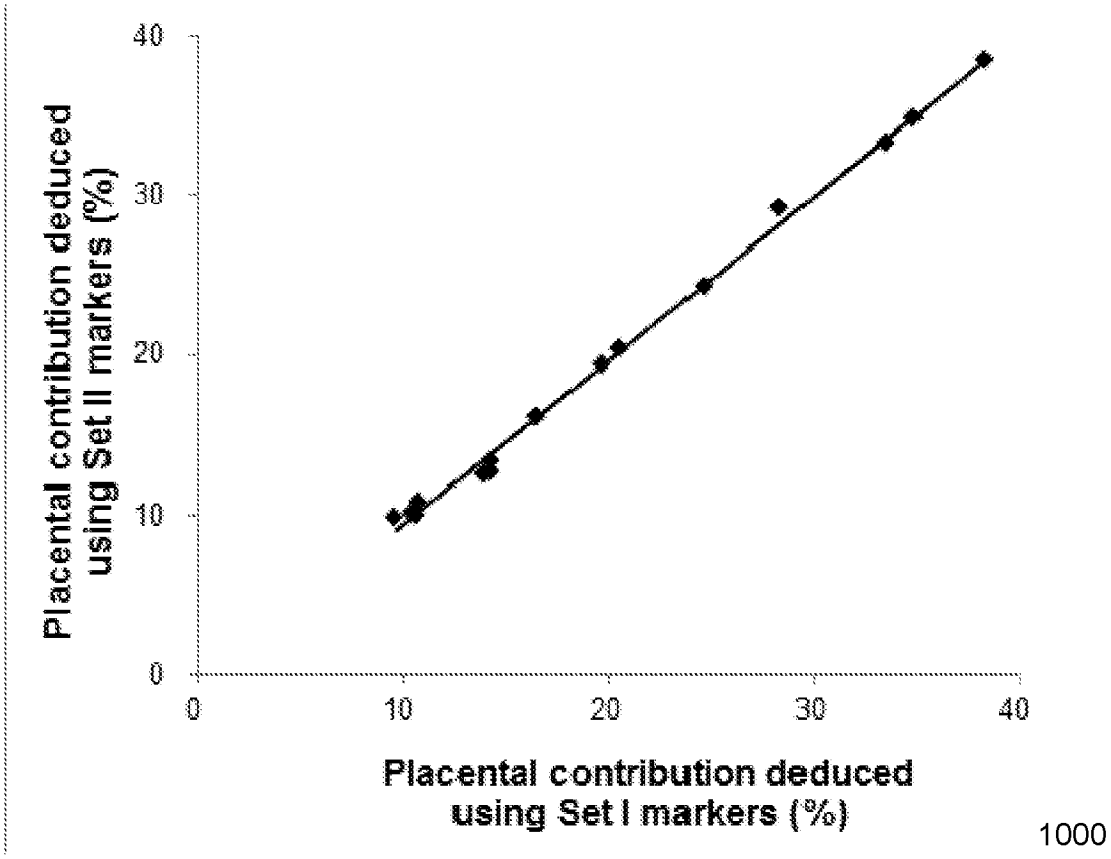


FIG. 10A

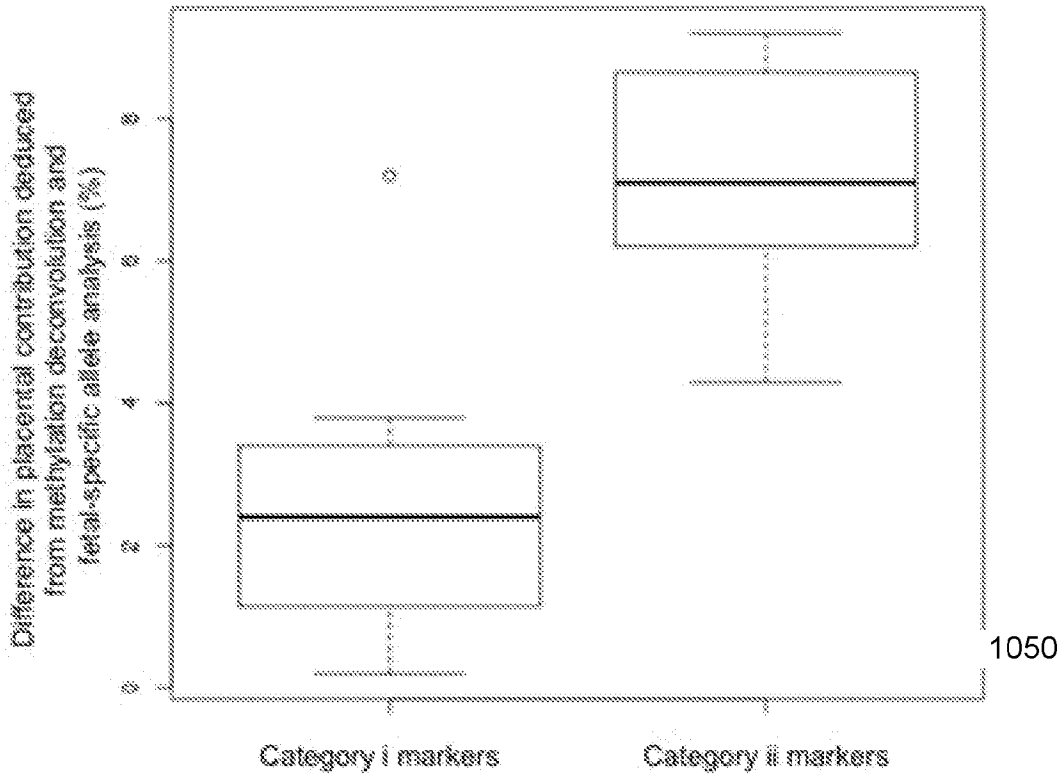


FIG. 10B

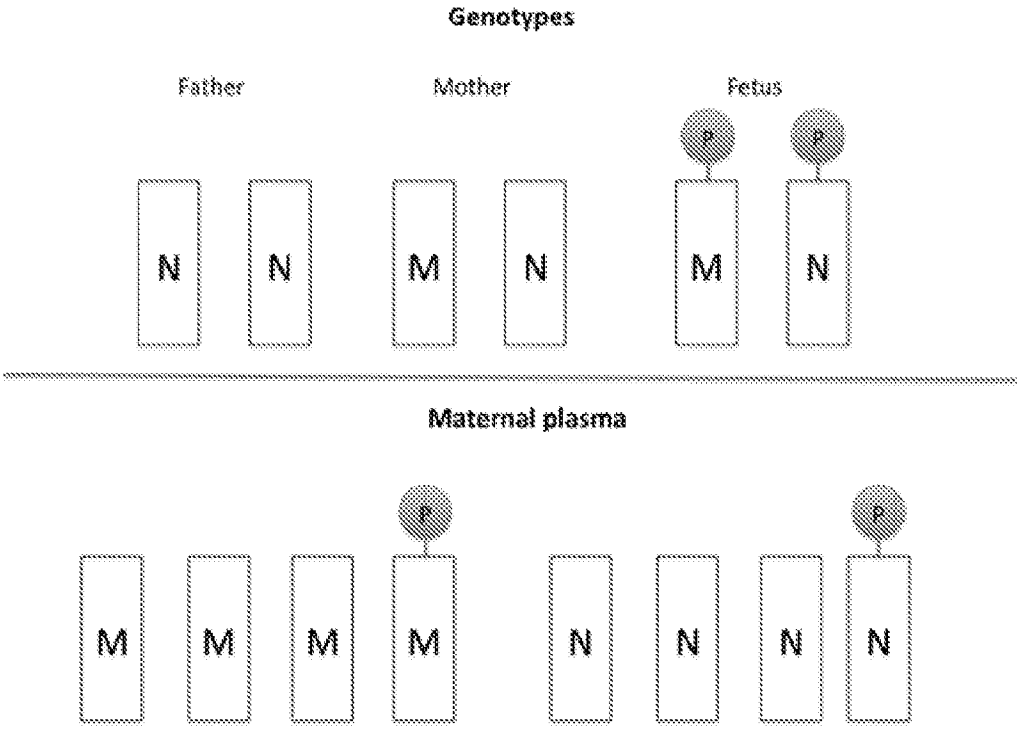


FIG. 11A

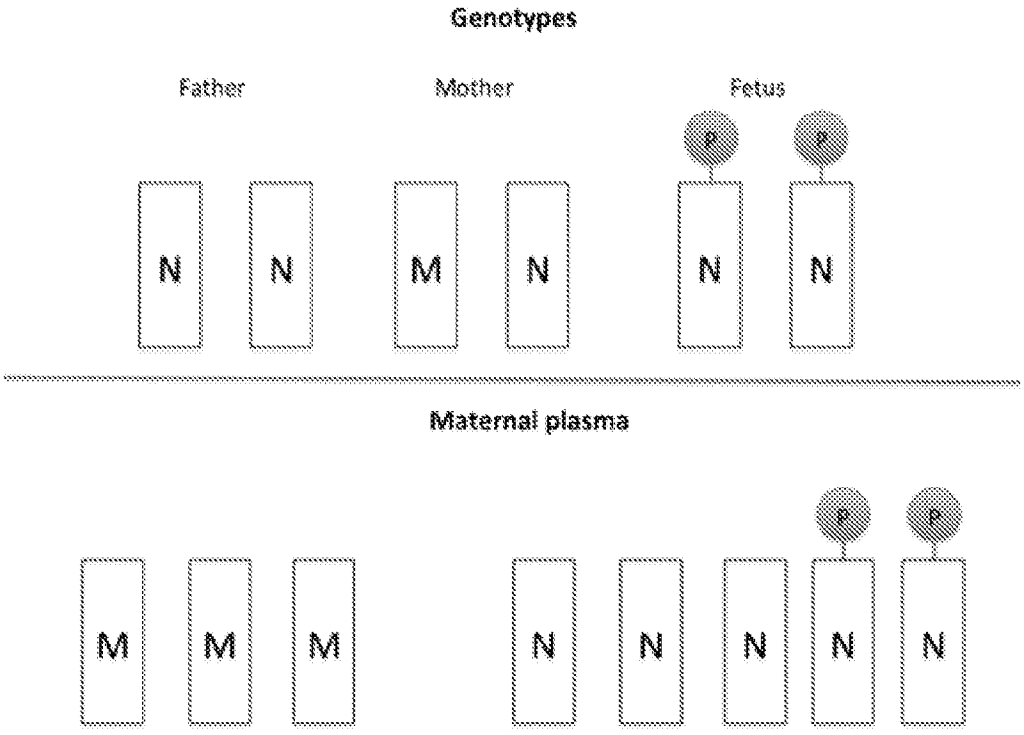


FIG. 11B

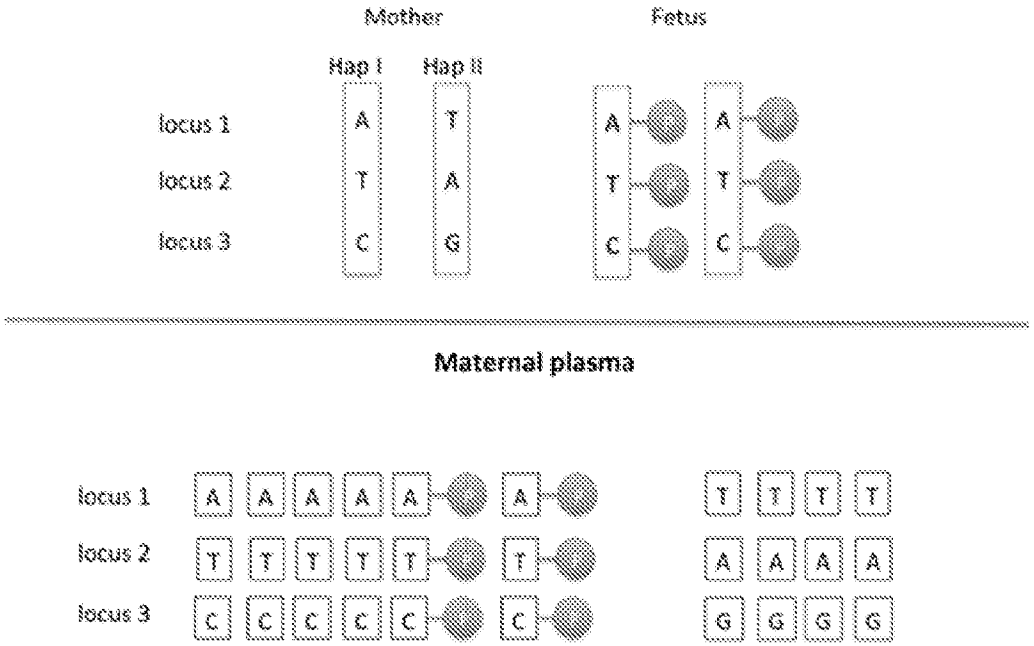


FIG. 12A

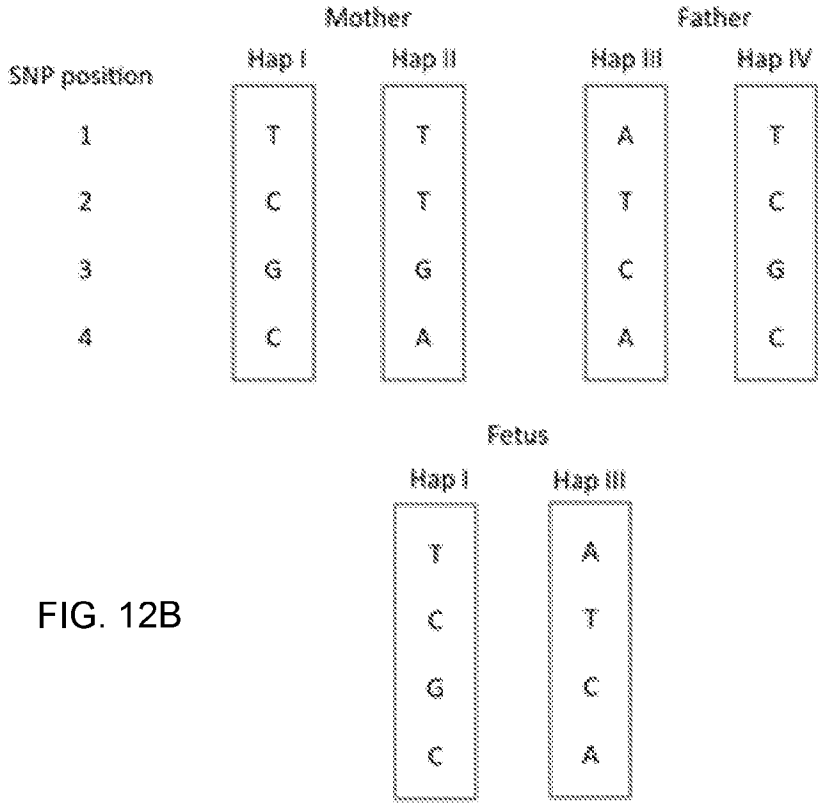


FIG. 12B

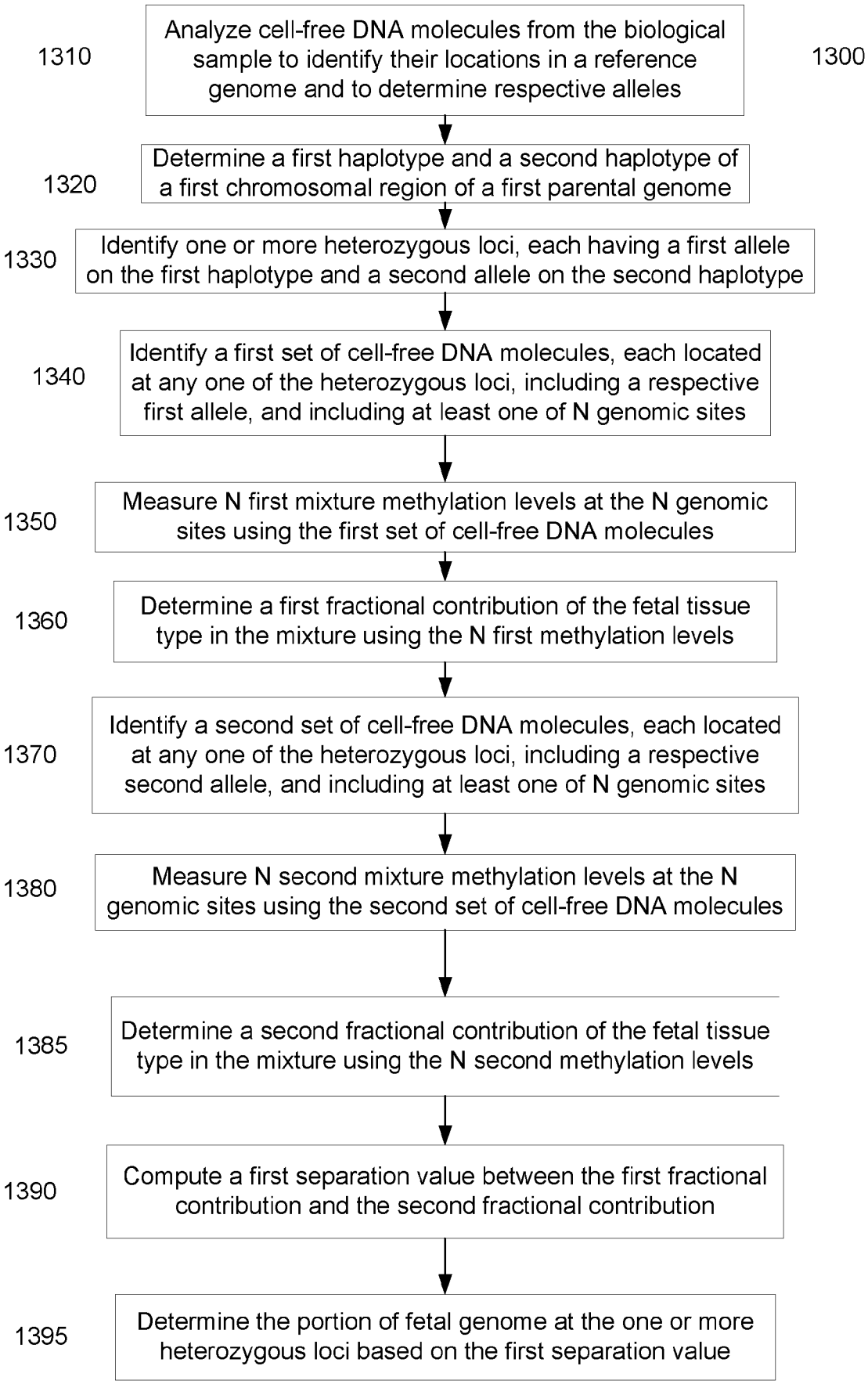

1300
1310
Analyze cell-free DNA molecules from the biological sample to identify their locations in a reference genome and to determine respective alleles
1320
Determine a first haplotype and a second haplotype of a first chromosomal region of a first parental genome
1330
Identify one or more heterozygous loci, each having a first allele on the first haplotype and a second allele on the second haplotype
1340
Identify a first set of cell-free DNA molecules, each located at any one of the heterozygous loci, including a respective first allele, and including at least one of N genomic sites
1350
Measure N first mixture methylation levels at the N genomic sites using the first set of cell-free DNA molecules
1360
Determine a first fractional contribution of the fetal tissue type in the mixture using the N first methylation levels
1370
Identify a second set of cell-free DNA molecules, each located at any one of the heterozygous loci, including a respective second allele, and including at least one of N genomic sites
1380
Measure N second mixture methylation levels at the N genomic sites using the second set of cell-free DNA molecules
1385
Determine a second fractional contribution of the fetal tissue type in the mixture using the N second methylation levels
1390
Compute a first separation value between the first fractional contribution and the second fractional contribution
1395
Determine the portion of fetal genome at the one or more heterozygous loci based on the first separation value

FIG. 13

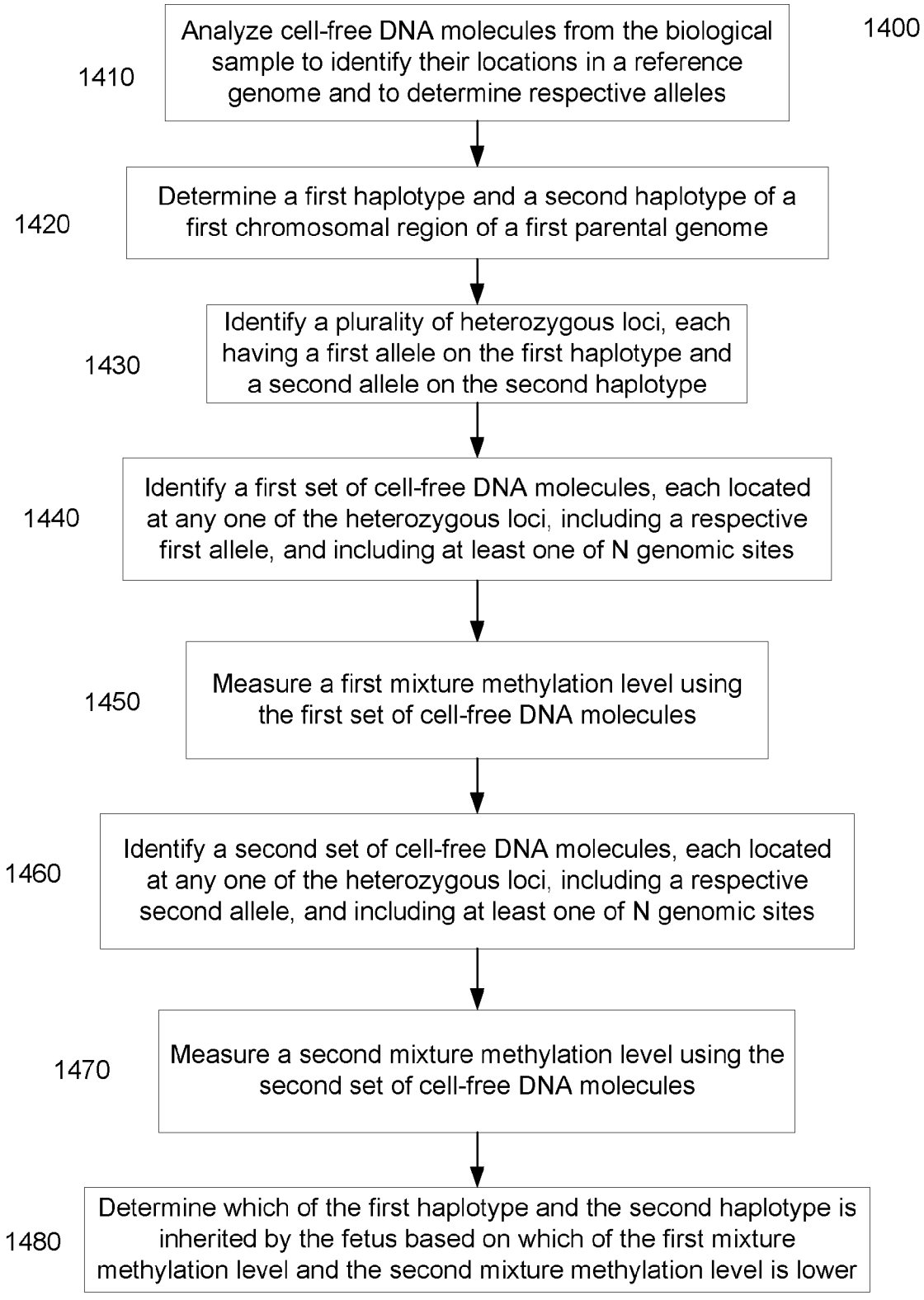

1400
1410
Analyze cell-free DNA molecules from the biological sample to identify their locations in a reference genome and to determine respective alleles
1420
Determine a first haplotype and a second haplotype of a first chromosomal region of a first parental genome
1430
Identify a plurality of heterozygous loci, each having a first allele on the first haplotype and a second allele on the second haplotype
1440
Identify a first set of cell-free DNA molecules, each located at any one of the heterozygous loci, including a respective first allele, and including at least one of N genomic sites
1450
Measure a first mixture methylation level using the first set of cell-free DNA molecules
1460
Identify a second set of cell-free DNA molecules, each located at any one of the heterozygous loci, including a respective second allele, and including at least one of N genomic sites
1470
Measure a second mixture methylation level using the second set of cell-free DNA molecules
1480
Determine which of the first haplotype and the second haplotype is inherited by the fetus based on which of the first mixture methylation level and the second mixture methylation level is lower

FIG. 14

FIG. 15

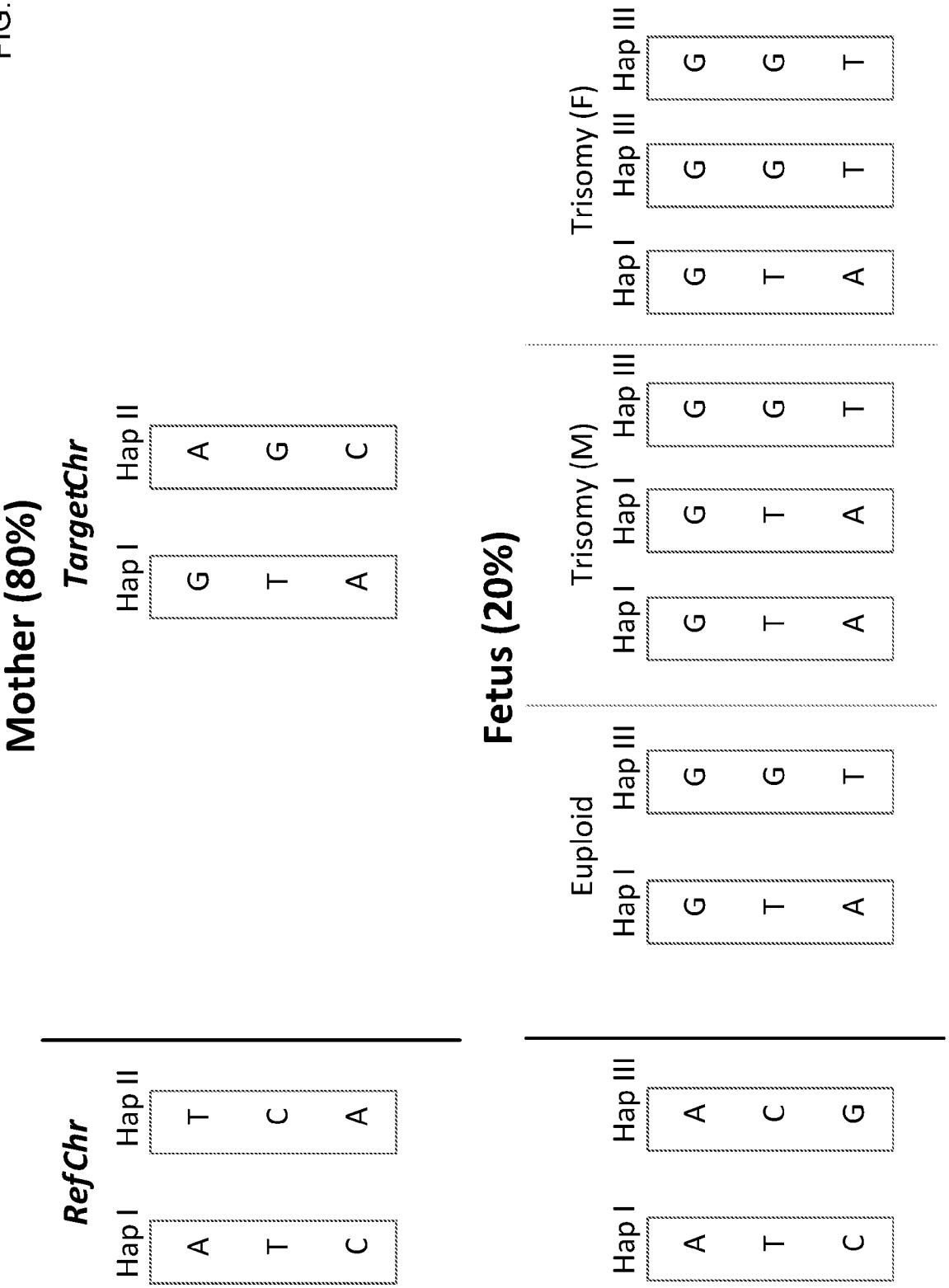
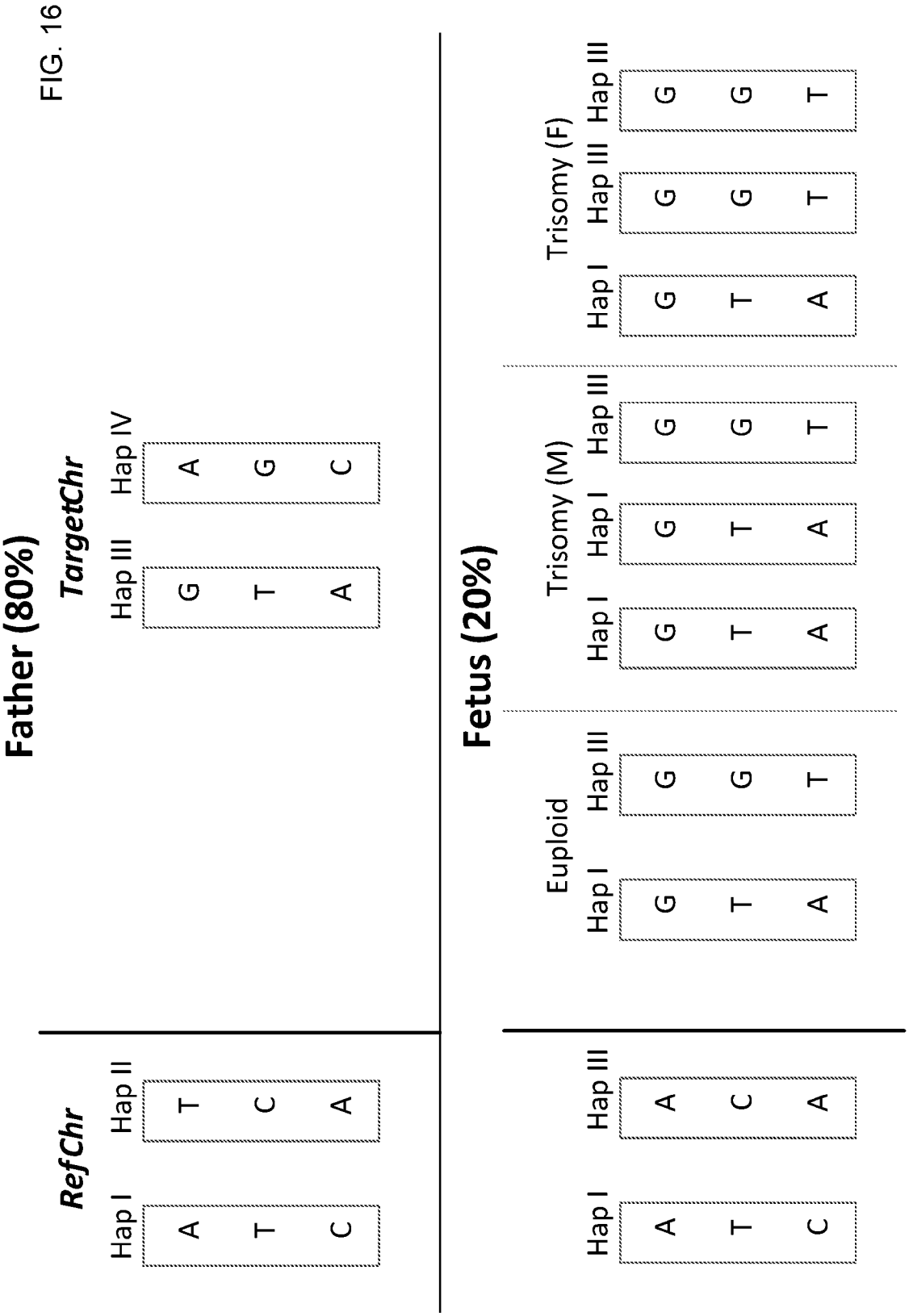

Mother (80%)
TargetChr
Hap I
Hap II
G
T
A
A
G
C
RefChr
Hap I
Hap II
A
T
C
T
C
A
Fetus (20%)
Euploid
Hap I
Hap III
G
T
A
G
G
T
Trisomy (M)
Hap I
Hap III
G
T
A
G
G
T
Trisomy (F)
Hap I
Hap III
G
T
A
G
G
T
Hap I
Hap III
A
T
C
A
C
G

FIG. 16


Father (80%)
RefChr
Hap I
A
T
C
Hap II
T
C
A
TargetChr
Hap III
G
T
A
Hap IV
A
G
C
Fetus (20%)
Hap I
A
T
C
Hap III
A
C
A
Euploid
Hap I
G
T
A
Hap III
G
G
T
Trisomy (M)
Hap I
G
T
A
Hap III
G
G
T
Trisomy (F)
Hap I
G
T
A
Hap III
G
G
T

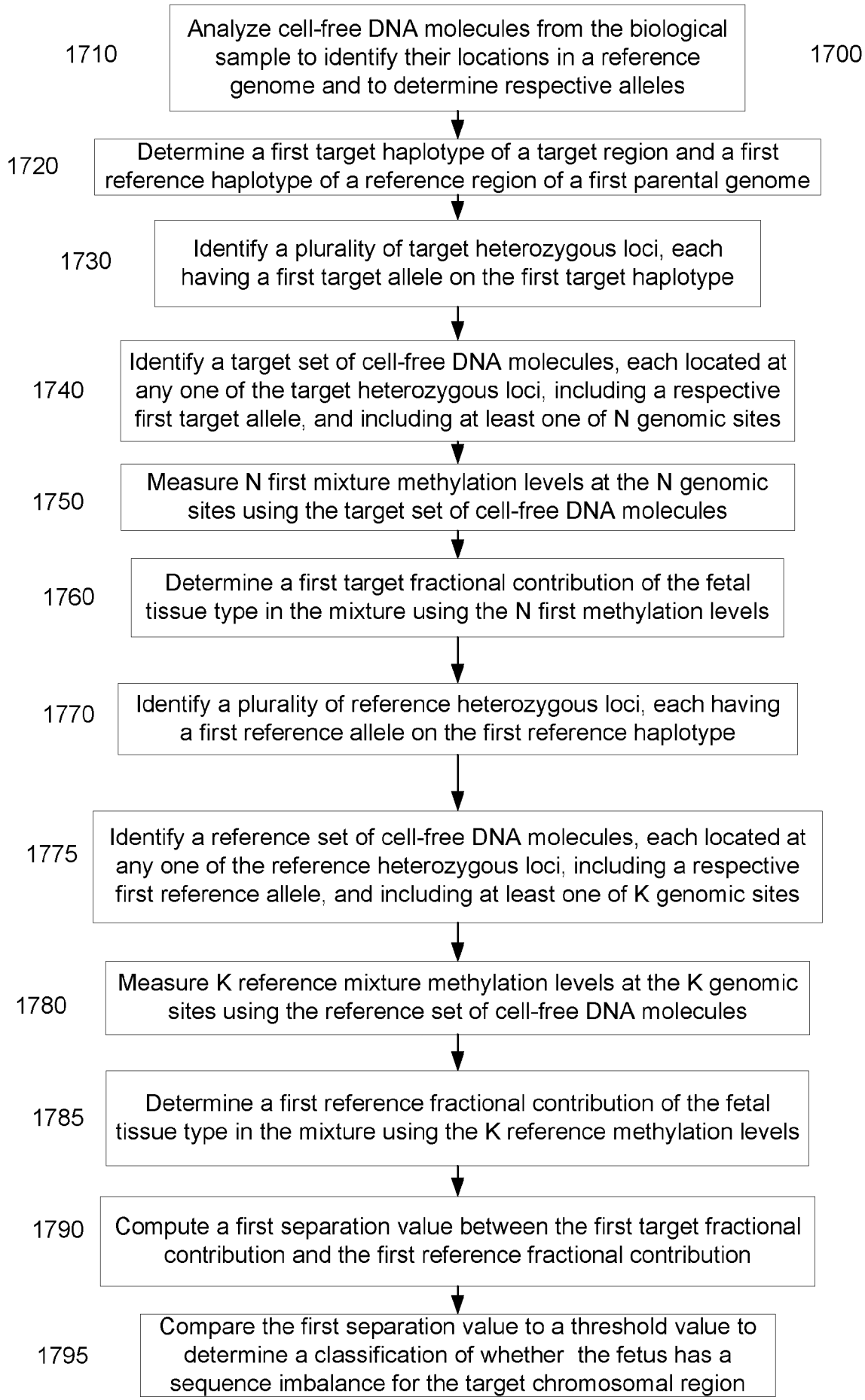

1700
1710
Analyze cell-free DNA molecules from the biological sample to identify their locations in a reference genome and to determine respective alleles
1720
Determine a first target haplotype of a target region and a first reference haplotype of a reference region of a first parental genome
1730
Identify a plurality of target heterozygous loci, each having a first target allele on the first target haplotype
1740
Identify a target set of cell-free DNA molecules, each located at any one of the target heterozygous loci, including a respective first target allele, and including at least one of N genomic sites
1750
Measure N first mixture methylation levels at the N genomic sites using the target set of cell-free DNA molecules
1760
Determine a first target fractional contribution of the fetal tissue type in the mixture using the N first methylation levels
1770
Identify a plurality of reference heterozygous loci, each having a first reference allele on the first reference haplotype
1775
Identify a reference set of cell-free DNA molecules, each located at any one of the reference heterozygous loci, including a respective first reference allele, and including at least one of K genomic sites
1780
Measure K reference mixture methylation levels at the K genomic sites using the reference set of cell-free DNA molecules
1785
Determine a first reference fractional contribution of the fetal tissue type in the mixture using the K reference methylation levels
1790
Compute a first separation value between the first target fractional contribution and the first reference fractional contribution
1795
Compare the first separation value to a threshold value to determine a classification of whether the fetus has a sequence imbalance for the target chromosomal region

FIG. 17

Donor
Recipient
Hap I
Hap II
Hap III
Hap IV
Locus 1
Locus 2
Locus 3
Locus 4
T
C
G
C
T
T
G
A
A
T
C
A
A
C
C
C
Plasma DNA deconvolution
Stable
Rejection
Tissues
Hap I
Hap II
Hap I
Hap II
Transplanted organ
Baseline level

FIG. 18

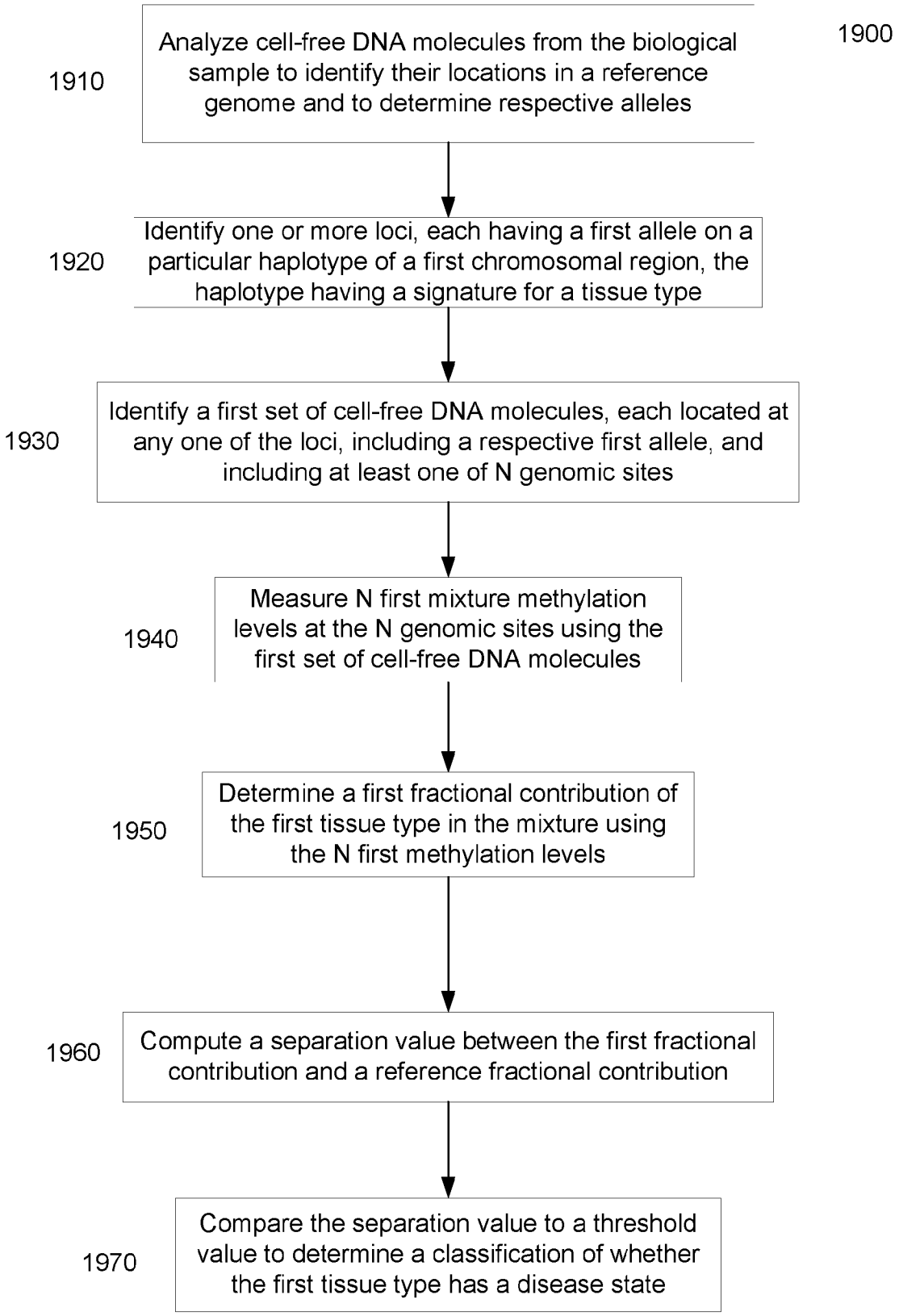

1900
1910
Analyze cell-free DNA molecules from the biological sample to identify their locations in a reference genome and to determine respective alleles
1920
Identify one or more loci, each having a first allele on a particular haplotype of a first chromosomal region, the haplotype having a signature for a tissue type
1930
Identify a first set of cell-free DNA molecules, each located at any one of the loci, including a respective first allele, and including at least one of N genomic sites
1940
Measure N first mixture methylation levels at the N genomic sites using the first set of cell-free DNA molecules
1950
Determine a first fractional contribution of the first tissue type in the mixture using the N first methylation levels
1960
Compute a separation value between the first fractional contribution and a reference fractional contribution
1970
Compare the separation value to a threshold value to determine a classification of whether the first tissue type has a disease state

FIG. 19

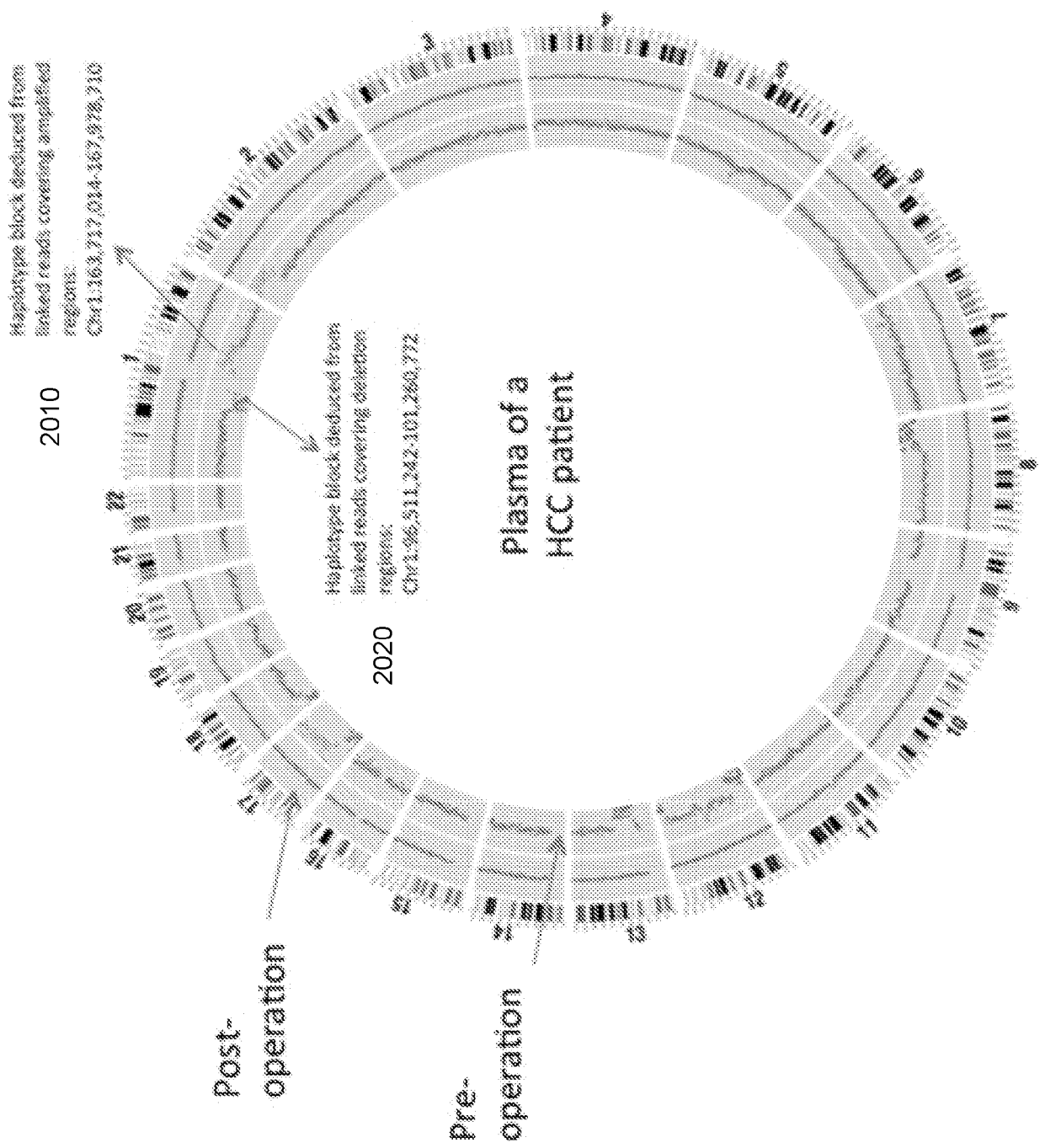

Haplotype block deduced from linked reads covering amplified regions:
Chr1:363,717,034-367,978,710
2010
Haplotype block deduced from linked reads covering deletion regions:
Chr1:96,311,242-101,260,772
2020
Plasma of a HCC patient
Post-operation
Pre-operation

FIG. 20

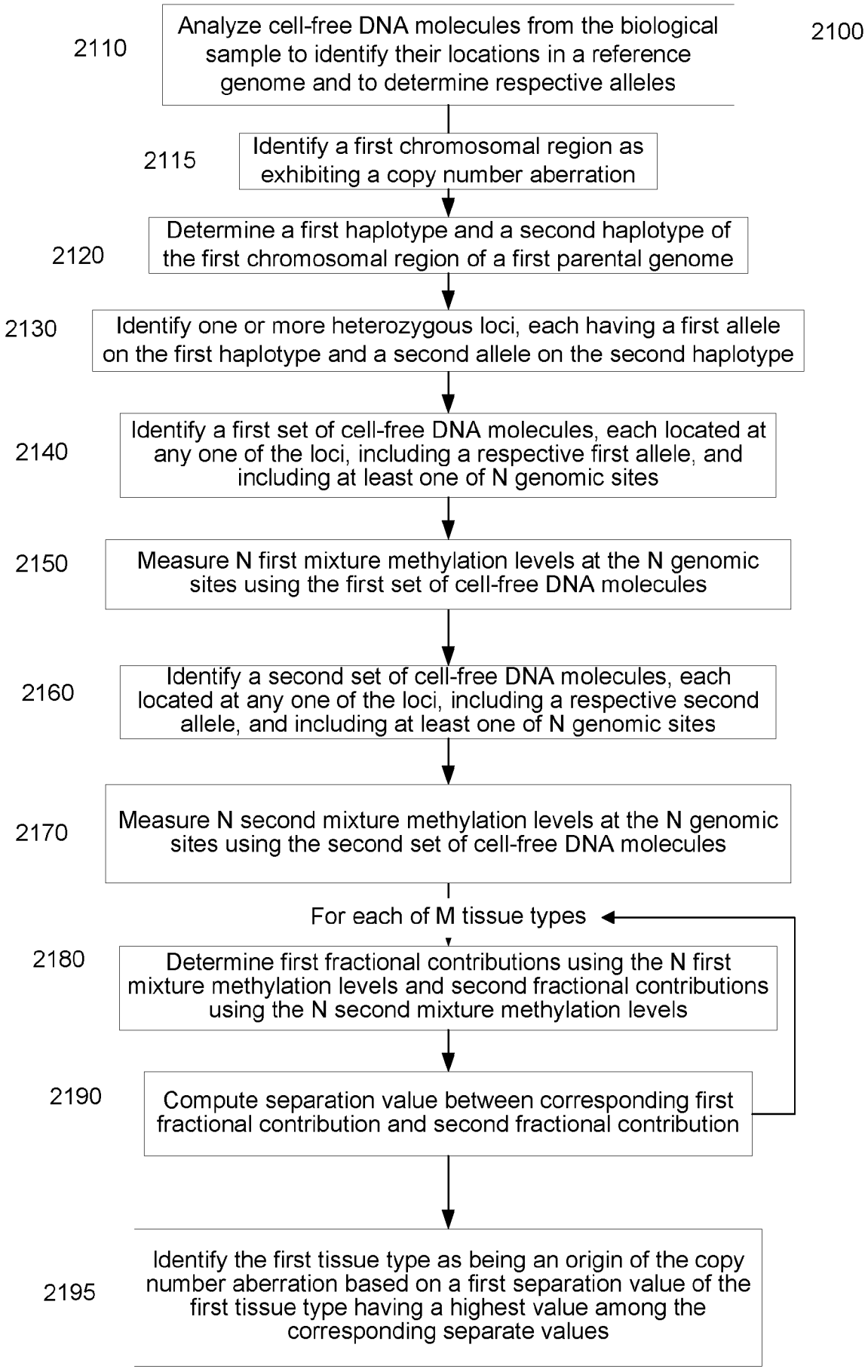

2100
2110
Analyze cell-free DNA molecules from the biological sample to identify their locations in a reference genome and to determine respective alleles
2115
Identify a first chromosomal region as exhibiting a copy number aberration
2120
Determine a first haplotype and a second haplotype of the first chromosomal region of a first parental genome
2130
Identify one or more heterozygous loci, each having a first allele on the first haplotype and a second allele on the second haplotype
2140
Identify a first set of cell-free DNA molecules, each located at any one of the loci, including a respective first allele, and including at least one of N genomic sites
2150
Measure N first mixture methylation levels at the N genomic sites using the first set of cell-free DNA molecules
2160
Identify a second set of cell-free DNA molecules, each located at any one of the loci, including a respective second allele, and including at least one of N genomic sites
2170
Measure N second mixture methylation levels at the N genomic sites using the second set of cell-free DNA molecules
For each of M tissue types
2180
Determine first fractional contributions using the N first mixture methylation levels and second fractional contributions using the N second mixture methylation levels
2190
Compute separation value between corresponding first fractional contribution and second fractional contribution
2195
Identify the first tissue type as being an origin of the copy number aberration based on a first separation value of the first tissue type having a highest value among the corresponding separate values

FIG. 21

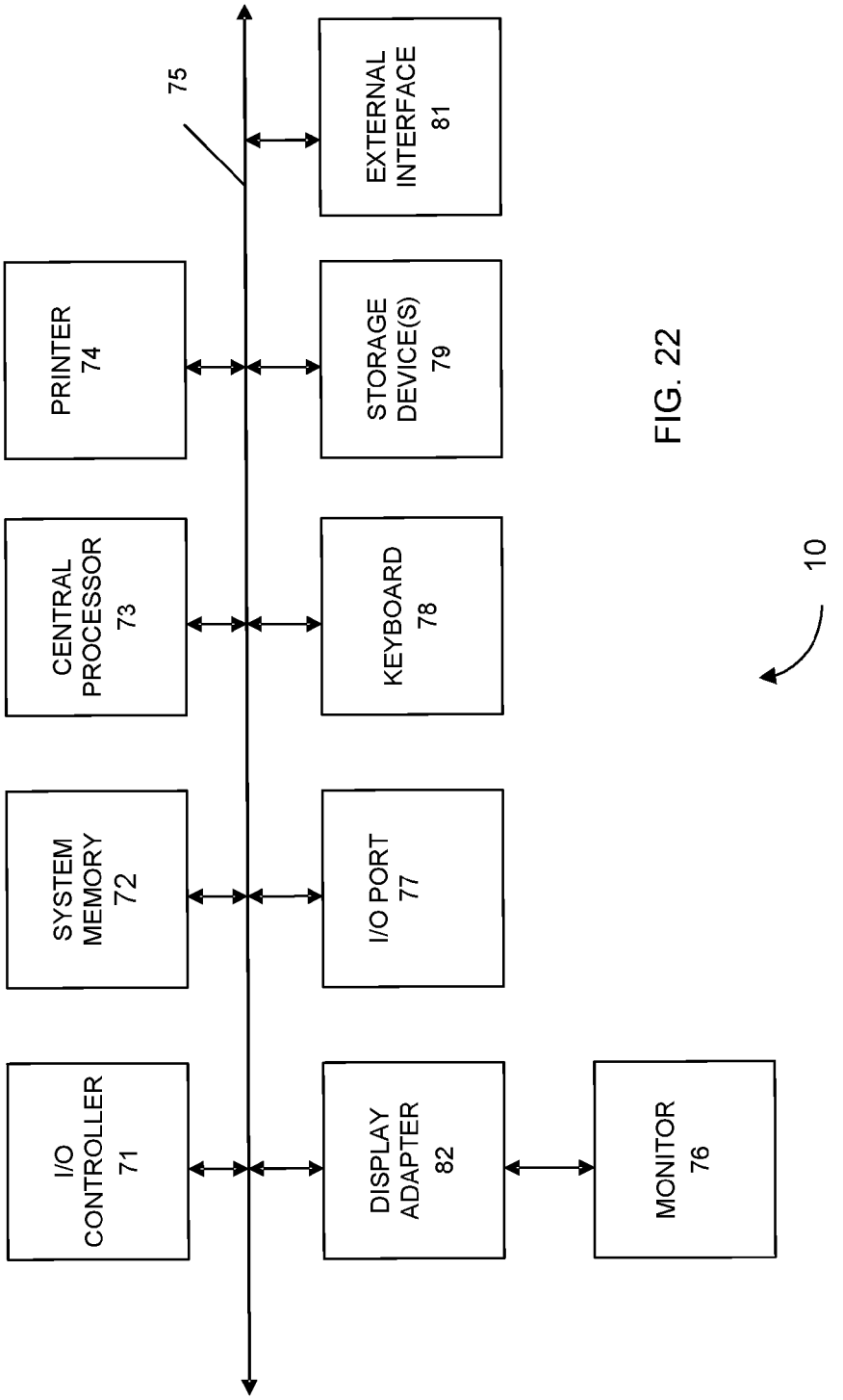

10
75
I/O CONTROLLER 71
SYSTEM MEMORY 72
CENTRAL PROCESSOR 73
PRINTER 74
DISPLAY ADAPTER 82
I/O PORT 77
KEYBOARD 78
STORAGE DEVICE(S) 79
MONITOR 76
EXTERNAL INTERFACE 81

FIG. 22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/090642

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI,SIPOABS,CNABS,HKABS,CNKI,Pubmed,Google scholar:fetal, fetus, allele, haplotype, chromosome, parental, methylation, cell-free, non-invasive, THE CHINESE UNIVERSITY OF HONG KONG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2014043763 A1 (UNIV HONG KONG CHINESEET AL.) 27 March 2014 (2014-03-27) see the whole document	1-65
A	WO 2011057094 A1 (UNIV HONG KONG CHINESEET AL.) 12 May 2011 (2011-05-12) see the whole document	1-65

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 October 2016	**19 October 2016**
Name and mailing address of the ISA/CN **STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088** **China**	Authorized officer **LI,Kangqi**
Facsimile No. **(86-10)62019451**	Telephone No. **(86-10)62411034**

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/090642

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO 2014043763 A1	27 March 2014	JP	2015536639	A	24 December 2015
		AU	2013317708	A1	09 April 2015
		HK	1207124	A1	22 January 2016
		KR	20150082228	A	15 July 2015
		MX	2015003571	A	22 June 2015
		CA	2884066	A1	27 March 2014
		EP	2898100	A1	29 July 2015
		TW	201418474	A	16 May 2014
		EA	201500327	A1	30 July 2015
		SG	11201501927V	A	29 April 2015
		HK	1206797	A1	15 January 2016
		EP	2898100	A4	11 May 2016
		IL	237495	A	30 April 2015
		CN	104781422	A	15 July 2015
		PH	12015500547	A1	04 May 2015
		IL	237495	D0	30 April 2015
WO 2011057094 A1	12 May 2011	IL	219521	A	26 February 2015
		CN	105779280	A	20 July 2016
		IL	219521	D0	28 June 2012
		EP	2496717	A1	12 September 2012
		AU	2010315037	B2	18 September 2014
		US	2013253844	A1	26 September 2013
		US	8467976	B2	18 June 2013
		AU	2010315037	A8	12 July 2012
		JP	5540105	B2	02 July 2014
		CA	2779695	A1	12 May 2011
		CA	2779695	C	24 May 2016
		TW	I458976	B	01 November 2014
		MX	2012005214	A	21 September 2012
		AU	2010315037	B9	23 April 2015
		TW	201122470	A	01 July 2011
		US	2013323731	A1	05 December 2013
		JP	2014193165	A	09 October 2014
		EA	201200690	A1	30 May 2013
		CN	102770558	B	06 April 2016
		AU	2010315037	A1	07 June 2012
		CN	102770558	A	07 November 2012
		US	2011105353	A1	05 May 2011
		JP	2013509884	A	21 March 2013

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 March 2005 (17.03.2005)

PCT

(10) International Publication Number
WO 2005/024053 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number:
PCT/AU2004/001196

(22) International Filing Date:
3 September 2004 (03.09.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2003904832 4 September 2003 (04.09.2003) AU

(71) Applicant *(for all designated States except US)*: **HUMAN GENETIC SIGNATURES PTY LTD** [AU/AU]; 63 Kananook Avenue, Bayview, NSW 2104 (AU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MILLAR, Douglas, Spencer** [AU/AU]; 8 Rowland Avenue, Revesby, NSW 2212 (AU). **MELKI, John, Robert** [AU/AU]; 7/170 Russell Avenue, Dolls Point, NSW 2219 (AU). **MIKLOS, George, L., Gabor** [AU/AU]; 81 Bynya Road, Palm Beach, NSW 2108 (AU).

(74) Agent: **ALLENS ARTHUR ROBINSON PATENT & TRADE MARKS ATTORNEYS**; The Chifley Tower, 2 Chifley Square, Sydney, NSW 2000 (AU).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACID DETECTION ASSAY

(57) Abstract: A method for determining the methylation status of a potential methylation site in genomic nucleic acid comprising treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5methyl-cytosine bases under conditions to form a. modified nucleic acid template containing a potential methylation site; providing a first clamp containing a first capture sequence complementary to a region flanking one side of the potential methylation site in the modified nucleic acid template, providing a second clamp containing a second capture sequence complementary to a region flanking the other side of the potential methylation site in the modified nucleic acid template; allowing the first clamp and the second clamp to hybridise to the modified nucleic acid template; ligating the hybridised first and second clamps to form a probe spanning the potential methylation site in the modified nucleic acid template; digesting the modified nucleic acid template to obtain the probe; and detecting the probe and determining the methylation status of the potential methylation site in the modified genomic nucleic acid.

WO 2005/024053 A1

NUCLEIC ACID DETECTION ASSAY

Technical Field

The invention relates to nucleic acid detection assays and in particular to improved oligonucleotide assays using a bisulphite clamp. The invention also relates to methods for distinguishing specific base sequences including 5-methyl cytosine bases in nucleic acids using these assays.

Background Art

A number of procedures are presently available for the detection of specific nucleic acid molecules. These procedures typically depend on sequence-dependent hybridisation between the target nucleic acid and nucleic acid probes which may range in length from short oligonucleotides (20 bases or less) to sequences of many kilobases (kb).

The most widely used method for amplification of specific sequences from within a population of nucleic acid sequences is that of polymerase chain reaction (PCR) (Dieffenbach C and Dveksler G eds. PCR Primer: A Laboratory Manual. Cold Spring Harbor Press, Plainview NY). In this amplification method, oligonucleotides, generally 15 to 30 nucleotides in length on complementary strands and at either end of the region to be amplified, are used to prime DNA synthesis on denatured single-stranded DNA. Successive cycles of denaturation, primer hybridisation and DNA strand synthesis using thermostable DNA polymerases allows exponential amplification of the sequences between the primers. RNA sequences can be amplified by first copying using reverse transcriptase to produce a cDNA copy. Amplified DNA fragments can be detected by a variety of means including gel electrophoresis, hybridisation with labelled probes, use of tagged primers that allow subsequent identification (eg. by an enzyme linked assay), use of fluorescently-tagged primers that give rise to a signal upon hybridisation with the target DNA (eg. Beacon and TaqMan systems).

As well as PCR, a variety of other techniques have been developed for detection and amplification of specific sequences. One example is the ligase chain reaction (Barany F Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc. Natl. Acad. Sci. USA 88:189-193 (1991)).

For direct detection, the target nucleic acid is most commonly separated on the basis of size by gel electrophoresis and transferred to a solid support prior to

hybridisation with a probe complementary to the target sequence (Southern and Northern blotting). The probe may be a natural nucleic acid or analogue such as peptide nucleic acid (PNA) or locked nucleic acid (LNA). The probe may be directly labelled (eg. with ^{32}P) or an indirect detection procedure may be used. Indirect procedures usually rely on incorporation into the probe of a "tag" such as biotin or digoxigenin and the probe is then detected by means such as enzyme-linked substrate conversion or chemiluminescence.

Another method for direct detection of nucleic acid that has been used widely is "sandwich" hybridisation. In this method, a capture probe is coupled to a solid support and the target nucleic acid, in solution, is hybridised with the bound probe. Unbound target nucleic acid is washed away and the bound nucleic acid is detected using a second probe that hybridises to the target sequences. Detection may use direct or indirect methods as outlined above. The "branched DNA" signal detection system is an example that uses the sandwich hybridization principle (Urdea MS et al. Branched DNA amplification multimers for the sensitive, direct detection of human hepatitis viruses. Nucleic Acids Symp Ser. 1991;(24):197-200).

A rapidly growing area that uses nucleic acid hybridisation for direct detection of nucleic acid sequences is that of DNA micro-arrays (Young RA Biomedical discovery with DNA arrays. Cell 102: 9-15 (2000); Watson A New tools. A new breed of high tech detectives. Science 289:850-854 (2000)). In this process, individual nucleic acid species, that may range from oligonucleotides to longer sequences such as cDNA clones, are fixed to a solid support in a grid pattern. A tagged or labelled nucleic acid population is then hybridised with the array and the level of hybridisation with each spot in the array quantified. Most commonly, radioactively- or fluorescently-labelled nucleic acids (eg. cDNAs) were used for hybridisation, though other detection systems were employed.

One problem associated with the use of micro-arrays for genomic typing analysis is that before individual sites can be analysed they usually have to be pre-amplified in some way. Most methods rely on PCR amplification of the target sequence. However, when the number of primer sets in the reaction mix is n target sequences, any $2n^2+n$ possible pairwise combination of probes may give rise to non-specific amplification products (Landegren and Nilsson. Locked on target: strategies for future gene diagnostics. Ann. Med. 1997; (29): 585-590). To address this technique, padlock probes have been developed. With these probes, non-specific reactions create linearised molecules whereas specific hybridization events lead to dimeric molecules which can be

distinguished from the linearised molecules by the use of exonucleases (Nilsson et al. Padlock probes: circularised oligonucleotides for local DNA detection. 1994; Science 92650; 2085-2088). However, for the specific detection of genomic variations typically four probes would be used for the detection of each individual polymorphic site leading to a very large number of probes to be generated for whole genome scanning.

Another recent technique, molecular inversion probes (MIP) has been demonstrated to produce a high level of multiplexing in a single tube. It has been reported that more that 1000 probes can be multiplexed in a single tube (Hardenbol et al Nature Biotechnology, 21: 673-678). However, both the padlock probe and MIP method require the synthesis of very long oligonucleotides >110 bp which cannot be achieved with most commercial distributors.

Currently, the method of choice to detect methylation changes in DNA, such as were found in the GSTP1 gene promoter in prostate cancer, are dependent on PCR amplification of such sequences after bisulphite modification of DNA. In bisulphite-treated DNA, cytosines are converted to uracils (and hence amplified as thymines during PCR) while methylated cytosines are non-reactive and remain as cytosines (Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL and Paul CL. A genomic sequencing protocol which yields a positive display of 5-methyl cytosine residues in individual DNA strands. PNAS 89: 1827-1831 (1992); Clark SJ, Harrison J, Paul CL and Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 22: 2990-2997 (1994)). Thus, after bisulphite treatment, DNA containing 5-methyl cytosine bases will be different in sequence from the corresponding unmethylated DNA. Primers may be chosen to amplify non-selectively a region of the genome of interest to determine its methylation status, or may be designed to selectively amplify sequences in which particular cytosines were methylated (Herman JG, Graff JR, Myohanen S, Nelkin BD and Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. PNAS 93:9821-9826 (1996)).

Alternative methods for detection of cytosine methylation include digestion with restriction enzymes whose cutting is blocked by site-specific DNA methylation, followed by Southern blotting and hybridisation probing for the region of interest. This approach is limited to circumstances where a significant proportion (generally >10%) of the DNA is methylated at the site and where there is sufficient DNA, about 1 to 5 μg, to allow for detection. Digestion with restriction enzymes whose cutting is blocked by site-specific DNA methylation is followed by PCR amplification using primers that flank the restriction enzyme site(s). This method can utilise smaller amounts of DNA but any lack of

complete enzyme digestion for reasons other than DNA methylation can lead to false positive signals.

The present inventors have now developed methods utilizing oligonucleotide clamps for the sensitive and specific detection of methylated nucleic acids which greatly reduce the problems associated with non-specific amplification of non-target sequences and increases the levels of multiplexing which can be carried out in an individual reaction tube. This makes the technique ideal for whole genome analysis of methylation patterns and also amenable to robotic manipulation.

Disclosure of Invention

In a first aspect, the present invention provides a method for determining the methylation status of a potential methylation site in genomic nucleic acid comprising:

(a) treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template containing a potential methylation site;

(b) providing a first clamp containing a first capture sequence complementary to a region flanking one side of the potential methylation site in the modified nucleic acid template;

(c) providing a second clamp containing a second capture sequence complementary to a region flanking the other side of the potential methylation site in the modified nucleic acid template;

(d) allowing the first clamp and the second clamp to hybridise to the modified nucleic acid template;

(e) ligating the hybridised first and second clamps to form a probe spanning the potential methylation site in the modified nucleic acid template;

(f) digesting the modified nucleic acid template; and

(g) detecting the probe and determining the methylation status of the potential methylation site in the modified genomic nucleic acid.

In a second aspect, the present invention provides a method for determining the methylation status of a potential methylation site in genomic nucleic acid comprising:

(a) treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template;

(b) providing a first clamp containing a first capture sequence complementary to a region flanking one side of the potential methylation site in the modified nucleic acid template;

(c) providing a second clamp containing a second capture sequence complementary to a region flanking the other side of the potential methylation site in the modified nucleic acid template;

(d) allowing the first clamp and the second clamp to hybridise to the modified nucleic acid template such that there is no complementary base(s) at the potential methylation site;

(e) causing a base or bases to be inserted between the first clamp and the second clamp;

(e) ligating the hybridised clamps to form a probe spanning the potential methylation site in the modified nucleic acid template;

(f) digesting the modified nucleic acid template; and

(g) detecting the probe and determining the methylation status of the potential methylation site in the modified genomic nucleic acid.

In a third aspect, the present invention provides a method for determining the methylation status of a potential methylation site on genomic nucleic acid comprising:

(a) treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template containing two complementary strands of nucleic acid containing a potential methylation site;

(b) providing a first clamp containing a first capture sequence and a second capture sequence, first capture sequence being complementary to a region flanking one side of the potential methylation site in the first strand of the modified nucleic acid template, and the second capture sequence being complementary to a region flanking one side of the potential methylation site in the second strand of the modified nucleic acid template;

(c) providing a second clamp containing a third capture sequence complementary to a region flanking the other side the potential methylation site in the first strand of the modified nucleic acid template;

(d) providing a third clamp containing a fourth capture sequence complementary to a region flanking the other side of the potential methylation site in the second strand of the modified nucleic acid template;

(e) allowing the first, second and third clamps to hybridise to the modified nucleic acid template;

(f) ligating the hybridised clamps to form a probe spanning the potential methylation site in the two strands of the modified nucleic acid template;

(g) digesting the modified nucleic acid template; and

(h) detecting the probe and determining the methylation status of the potential methylation sites in the modified genomic nucleic acid.

In a fourth aspect, the present invention provides a method for determining the methylation status of a potential methylation site on genomic nucleic acid comprising:

(a) treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template containing two complementary strands of nucleic acid containing a potential methylation site;

(b) providing a first clamp containing a first capture sequence and a second capture sequence, the first capture sequence being complementary to a region flanking one side of the potential methylation site in the first strand of the modified nucleic acid template, and the second capture sequence being complementary to a region flanking one side of the potential methylation site in the second strand of the modified nucleic acid template;

(c) providing a second clamp containing a third capture sequence complementary to a region flanking the other side the potential methylation site in the first strand of the modified nucleic acid template;

(d) providing a third clamp containing a fourth capture sequence complementary to a region flanking the other side of the potential methylation site in the second strand of the modified nucleic acid template;

(e) allowing the first, second and third clamps to hybridise to the modified nucleic acid template such that there is no complementary base at the potential methylation site of at least one of the complementary strands of the modified nucleic acid template;

(f) causing at least one base to be inserted between the first clamp and the second clamp;

(g) ligating the hybridised clamps to form a linear probe spanning the potential methylation site in the two strands of the modified nucleic acid template;

(h) digesting the modified nucleic acid template; and

(h) detecting the linear probe and determining the methylation status of the potential methylation sites in the modified genomic nucleic acid.

In a fifth aspect, the present invention provides a method for determining the methylation status of a potential methylation site on genomic nucleic acid comprising:

(a) treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template containing two complementary strands of nucleic acid containing a potential methylation site;

(b) providing a first clamp containing a first capture sequence and a second capture sequence, first capture sequence being complementary to a region flanking one side of the potential methylation site in the first strand of the modified nucleic acid template, and the second capture sequence being complementary to a region flanking one side of the potential methylation site in the second strand of the modified nucleic acid template;

(c) providing a second clamp containing a third capture sequence and a fourth capture sequence, the third capture sequence being complementary to a region flanking the other side the potential methylation site in the first strand of the modified nucleic acid template, and the fourth capture sequence being complementary to a region flanking the other side of the potential methylation site in the second strand of the modified nucleic acid template;

(d) allowing the first clamp and the second clamp to hybridise to the two strands of the modified nucleic acid template;

(e) ligating the hybridised clamps to form a circular probe spanning the potential methylation site in the two strands of the modified nucleic acid template;

(f) digesting the modified nucleic acid template; and

(g) detecting the circular probe and determining the methylation status of the potential methylation site in the modified genomic nucleic acid.

In a sixth aspect, the present invention provides a method for determining the methylation status of a potential methylation site on genomic nucleic acid comprising:

(a) treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template containing two complementary strands of nucleic acid containing a potential methylation site;

(b) providing a first clamp containing a first capture sequence and a second capture sequence, the first capture sequence being complementary to a region flanking one side of the potential methylation site in the first strand of the modified nucleic acid template, and the second capture sequence being complementary to a region flanking one side of the potential methylation site in the second strand of the modified nucleic acid template;

(c) providing a second clamp containing a third capture sequence and a fourth capture sequence, the third capture sequence being complementary to a region flanking the other side the potential methylation site in the first strand of the modified nucleic acid template, and the fourth capture sequence complementary to a region flanking the other side of the potential methylation site in the second strand of the modified nucleic acid template;

(d) allowing the first clamp and the second clamp to hybridise to the two strands of the modified nucleic acid template such that there is no complementary base at the potential methylation site of at least one of the complementary strands of the modified nucleic acid template;

(e) causing at least one base to be inserted between the first clamp and the second clamp;

(f) ligating the hybridised clamps to form a circular probe spanning the potential methylation site in the two strands of the modified nucleic acid template;

(g) digesting the modified nucleic acid template; and

(h) detecting the circular probe and determining the methylation status of the potential methylation site in the modified genomic nucleic acid.

The genomic nucleic acid can be DNA or RNA. Preferably, the genomic nucleic acid is DNA.

Preferably, the potential methylation site is cytosine (C) flanked 5' by a guanine (G), termed in the art as CpG doublet.

The modifying agent is preferably selected from bisulfite, acetate or citrate. More preferably, the agent is sodium bisulfite, a reagent, which in the presence of water, modifies cytosine into uracil.

Sodium bisulfite ($NaHSO_3$) reacts readily with the 5,6-double bond of cytosine to form a sulfonated cytosine reaction intermediate which is susceptible to deamination, and in the presence of water gives rise to a uracil sulfite. If necessary, the sulfite group can be removed under mild alkaline conditions, resulting in the formation of uracil. Thus, potentially all cytosines will be converted to uracils. Any methylated cytosines, however, cannot be converted by the modifying reagent due to protection by methylation.

In one preferred form, at least one of the first, second or third clamps contains a universal primer which allows amplification of the probe after the ligation step.

In another preferred form, at least one of the first, second or third clamps contains a capture site which allows capture of the probe.

In another preferred form, at least one of the first, second or third clamps contains a cleavable site. Preferably, the cleavable site is nuclease restriction site or a uracil base.

In one preferred form, the complementary sequence(s) of the first or second clamps spans the potential methylation site.

In another preferred form, the complementary sequence(s) of the first, second or third clamps does not span the potential methylation site.

The clamps are preferably ligated by the use of a suitable enzyme capable of ligating single stranded DNA. Examples of suitable ligases include, but not limited to, Ampligase (Epicentre various cat#'s),T4 DNA ligase (NEB cat# M0202), Thermus aquaticus DNA ligase (NEB cat# M0208), DNA ligase (E.coli, NAD) (NEB cat# M0205. It will be appreciated, however, that other ligases would also be suitable for the present invention.

In the fifth and sixth aspects, the clamps are preferably ligated at or near the first and second capture sequences and at or near the third and fourth capture sequences.

When one or more bases are inserted between the clamps, preferably a DNA polymerase is used. Examples of suitable polymerases include, but not limited to, Taq polymerase Stoffel fragment, Taq polymerase, Advantage DNA polymerase, AmpliTaq,

Amplitaq Gold, Titanium Taq polymerase, KlenTaq DNA polymerase, Platinum Taq polymersae, Accuprime Taq polymerase. DNA polymerase1, T4 DNA polymerase, T7 DNA polymerase, Klenow large fragment, Klenow exo- DNA polymerase, BstB polymerase. Or any suitable DNA polymerase. It will be appreciated, however, that other polymerases would also be suitable for the present. In this form, preferably the base insertion reaction is carried out in the presence of a single nucleotide type and each nucleotide type is reacted separately. Thus, the insertion of a particular base will be indicative of the identity of complementary base on the modified DNA template.

Preferably, the modified DNA template is digested by the use of Uracil N DNA Glycosylase (UDG or UNG). As the treated DNA template will contain uracils, it is possible to remove the template by treatment with UDG. It will be appreciated, however, that any other digestion method that does not substantially digest the probe would also be suitable for the present invention. An advantage of removal of the nucleic acid template is that any subsequent amplification of the probe will have much less background and allow more accurate amplification and subsequent detection. Furthermore the removal of the nucleic acid template reduces the amount of artefactual products that may non-specifically form.

Suitable methods to detect the probe and determine the methylation status of the potential methylation site in the modified genomic nucleic acid include, but not limited to, incorporation of the missing base(s) between the first and second clamps, specific amplification of the probe. Numerous possible detection systems exist to determine the epigenetic status of the desired sample. Detection systems include but not limited to:

I. Hybridization of appropriately labelled amplified nucleic acid to a micro-array type device which could select for 10->60,000 individual components. The arrays could be composed of either intercalating nucleic acids (INAs, see WO 03/051901), PNAs or nucleotide or modified nucleotides arrays onto any suitable solid surface such as glass, plastic, mica, nylon , bead, magnetic bead, fluorescent bead or membrane.

II. Southern blot type detection systems.

III. Real-Time PCR quantitation of specific or multiple genomic amplified fragments or any variation such as molecular beacons, scorpions and the like.

IV. Detection systems using fluorescent beads, enzyme conjugates, radioactive beads, etc.

V. Any other amplification system such as ligase chain reaction, etc.

VI. Isothermal nucleic acid amplification technologies such as strand displacement amplification (SDA), rolling circle amplification.

Prior to the present invention, detection of a site of interest in genomic nucleic acid has not been achieved using bisulphite treatment, specific hybridisation of two 'flanking' clamps, optionally base insertion between the clamps, ligation of the two clamps, followed by removal of the template nucleic acid. The treated nucleic acid acts as a scaffold or template for aligning of the first and second clamps and is then removed. In contrast, most prior art assays require capture and the retention of the nucleic acid template.

One advantage of the present invention is that it does not necessarily require amplification of the genomic nucleic acid for the method to be used.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed in Australia prior to development of the present invention.

In order that the present invention may be more clearly understood, preferred forms will be described with reference to the following drawings and examples.

Brief Description of the Drawings

Figure 1 shows effect of Exonuclease III concentration and Ligase concentration on Clamp specificity (according to sixth aspect).

Figure 2 shows detection of methylated, unmethylated and a mixture of 50% methylated + 50% unmethylated DNA (according to sixth aspect).

Figure 3 shows effect of polymerase on the clamp reaction (according to sixth aspect).

Figure 4 shows effect of Polymerase/Ligase concentration (according to sixth aspect).

Figure 5 shows sensitivity of the Clamp reaction (according to third aspect) using clamp primers directed against methylated DNA sequences.

Figure 6 shows clamp reaction on genomic DNA (known to be unmethylated at the site of interrogation) (according to sixth aspect).

Figure 7 shows clamp reaction (according to second aspect).

Figure 8 shows examples of the first aspect of the present invention.

Figure 9 shows examples of the second aspect of the present invention.

Figure 10 shows examples of the third aspect of the present invention.

Figure 11 shows examples of the fourth aspect of the present invention.

Figure 12 shows examples of the fifth aspect of the present invention.

Figure 13 shows examples of the sixth aspect of the present invention.

Figure 14 is a schematic representation of the Bisulphite Clamp method according to the present invention.

Figure 15 shows a ligation dependant reaction.

Figure 16 shows an array tag detection system.

Mode(s) for Carrying Out the Invention

MATERIALS AND METHODS

Clamps

Each clamp can be an oligonucleotide or oligonucleotide analogue selected from DNA, RNA, locked nucleic acid (LNA), peptide nucleic acid (PNA), MNA, altritol nucleic acid (ANA), hexitol nucleic acid (HNA), intercalating nucleic acid (INA), cyclohexanyl nucleic acid (CNA) and mixtures thereof and hybrids thereof, as well as phosphorous atom modifications thereof, such as but not limited to phosphorothioates, methyl phospholates, phosphoramidites, phosphorodithiates, phosphoroselenoates, phosphotriesters and phosphoboranoates. Non-naturally occurring nucleotides include, but not limited to the nucleotides comprised within DNA, RNA, PNA, INA, HNA, MNA, ANA, LNA, CNA, CeNA, TNA, (2'-NH)-TNA, (3'-NH)-TNA, α-L-Ribo-LNA, α-L-Xylo-LNA, β-D-Xylo-LNA, α-D-Ribo-LNA, [3.2.1]-LNA, Bicyclo-DNA, 6-Amino-Bicyclo-DNA, 5-epi-

Bicyclo-DNA, α-Bicyclo-DNA, Tricyclo-DNA, Bicyclo[4.3.0]-DNA, Bicyclo[3.2.1]-DNA, Bicyclo[4.3,0]amide-DNA, β-D-Ribopyranosyl-NA, α-L-Lyxopyranosyl-NA, 2'-R-RNA, α-L-RNA or α-D-RNA, β-D-RNA. In addition non-phosphorous containing compounds may be used for linking to nucleotides such as but not limited to methyliminomethyl, formacetate, thioformacetate and linking groups comprising amides. In particular nucleic acids and nucleic acid analogues may comprise one or more intercalator pseudonucleotides.

Preferably, the clamps are DNA oligonucleotides.

Genomic nucleic acid

Genomic nucleic acid can be DNA or RNA obtained from plants, animals, microorganisms such as bacteria, fungi yeasts and viruses. Preferably, the nucleic acid is DNA, more preferably genomic DNA from an animal or human, or nucleic acid of an infectious agent of animal or human cells.

Bisulphite treatment of DNA

An exemplary protocol for effective bisulphite treatment of nucleic acid is set out below. The protocol results in retaining substantially all DNA treated. This method is also referred to herein as the Human Genetic Signatures (HGS) method. It will be appreciated that the volumes or amounts of sample or reagents can be varied.

Preferred method for bisulphite treatment can be found in US 10/428310 or PCT/AU2004/000549 incorporated herein by reference.

To 2 µg of DNA, which can be pre-digested with suitable restriction enzymes if so desired, 2 µl (1/10 volume) of 3 M NaOH (6g in 50 ml water, freshly made) was added in a final volume of 20 µl. This step denatures the double stranded DNA molecules into a single stranded form, since the bisulphite reagent preferably reacts with single stranded molecules. The mixture was incubated at 37°C for 15 minutes. Incubation at temperatures above room temperature can be used to improve the efficiency of denaturation.

After the incubation, 208 µl 2 M Sodium Metabisulphite (7.6 g in 20 ml water with 416 ml 10 N NaOH; BDH AnalaR #10356.4D; freshly made) and 12 µl of 10 mM Quinol (0.055 g in 50 ml water, BDH AnalR #103122E; freshly made) were added in succession. Quinol is a reducing agent and helps to reduce oxidation of the reagents. Other reducing

agents can also be used, for example, dithiothreitol (DTT), mercaptoethanol, quinone (hydroquinone), or other suitable reducing agents. The sample was overlaid with 200 µl of mineral oil. The overlaying of mineral oil prevents evaporation and oxidation of the reagents but is not essential. The sample was then incubated overnight at 55°C. Alternatively the samples can be cycled in a thermal cycler as follows: incubate for about 4 hours or overnight as follows: Step 1, 55°C / 2 hr cycled in PCR machine; Step 2, 95°C / 2 min. Step 1 can be performed at any temperature from about 37°C to about 90°C and can vary in length from 5 minutes to 8 hours. Step 2 can be performed at any temperature from about 70°C to about 99°C and can vary in length from about 1 second to 60 minutes, or longer.

After the treatment with Sodium Metabisulphite, the oil was removed, and 1 µl tRNA (20 mg/ml) or 2 µl glycogen were added if the DNA concentration was low. These additives are optional and can be used to improve the yield of DNA obtained by co-precitpitating with the target DNA especially when the DNA is present at low concentrations. The use of additives as carrier for more efficient precipitation of nucleic acids is generally desired when the amount nucleic acid is <0.5 µg.

An isopropanol cleanup treatment was performed as follows: 800 µl of water were added to the sample, mixed and then 1 ml isopropanol was added. The water or buffer reduces the concentration of the bisulphite salt in the reaction vessel to a level at which the salt will not precipitate along with the target nucleic acid of interest. The dilution is generally about 1/4 to 1/1000 so long as the salt concentration is diluted below a desired range, as disclosed herein.

The sample was mixed again and left at 4°C for a minimum of 5 minutes. The sample was spun in a microfuge for 10-15 minutes and the pellet was washed 2x with 70% ETOH, vortexing each time. This washing treatment removes any residual salts that precipitated with the nucleic acids.

The pellet was allowed to dry and then resuspended in a suitable volume of T/E (10 mM Tris/0.1 mM EDTA) pH 7.0-12.5 such as 50 µl. Buffer at pH 10.5 has been found to be particularly effective. The sample was incubated at 37°C to 95°C for 1 min to 96 hr, as needed to suspend the nucleic acids.

Addition of polymerase and ligase

The polymerase will add the correct base i.e. G (complementary to methylated) or A (complementary to unmethylated).

The ligase will then ligate the G or A base and a circle will be formed.

The bisulphite treated DNA is then degraded with Uracil N DNA glycosylase which also linearises the circle.

The Linear DNA is then amplified with a pair of common primers.

The DNA is then detected on a micro-array platform using a unique DNA sequence (Zip code) present in the clamp primers (NB not detected in traditional way as the Zip code only tells the complex where to go spatially on the array not by hybridization to the complementary strand of the bisulphite treated DNA).

NB Primers 1 and 3 are identical in sequence except for the inclusion of a different zip code (Blue and yellow bars) and primer 1 having a terminal base G which detects the presence of methylation while primer 3 has a terminal base A which detects the lack of methylation. This enables the products to be spatially separated on a micro-array (see Figure 15). This means that the presence of a C or T can be detected simultaneously in the same reaction vessel. Thus, thousands of different genomic loci can be interrogated in the one reaction vessel and separated on the basis of their unique zip codes.

After the clamps have been ligated and amplified they are then directed to specific sites on a micro-array devise predetermined by the internal Zip-code contained within each bisulphite clamp primer set (Figure 16).

RESULTS

Effect of Exonuclease III concentration and ligase concentration on clamp specificity

Experiment was carried out using methylated DNA sequences. Figure 1 shows the effect of ligase enzyme concentration and Exonuclease III concentration of the performance of the clamp reaction. As can be seen from the figure if too much exonuclease enzyme is used (10 Units), the reaction is not as efficient. Conversely if too little enzyme is used (2 Units) then this decreases the specificity of the reaction resulting in non-specific amplification occurring in he A, T and C tracks (see Figure 1A). The upper band seen if some lanes (indicated by the star) is a dimer of the target band as a results of the polymerase copying the circle twice.

In addition increasing the ligase concentration from 5U per reaction to 10U per reaction reduced the specificity of the method again resulting in non-specific detection in

the A, T and C tracks (see figure 1b). The bolded letters indicate the optimal exonuclease concentrations. One nanogram (ng) of a methylated oligonucleotide was used as a target. Ten ng of each clamp was used in the reaction, together with 0.1 units of Stoffel DNA polymerase. Ampligase enzyme was used at 5 and 10 units. The arrow indicates the correct size of the PCR products.

In detail, the following was added to a 0.5 µl thin-walled PCR tube:

10x Ampligase buffer	1.0 µl
Clamp#1 (10 ng/µl)	1.0 µl
Clamp#2 (10 ng/µl)	1.0 µl
Stoffel Polymerase	0.1 µl
Ampligase	1.0 µl or 2 µl
dG/dA/dT/dC (1 mM)	1.0 µl (dG/dA/dT/dC are added to separate tubes)
water	3.9 µl or 2.9 µl

Reactions were incubated as follows:

20°C 4 min, 95°C 2 min, 50°C 15 min, 55°C 5 min (with a 0.05°C/sec ramp time), 50°C 1 min.

Exonuclease treatment:

	10U ExoIII/2U Exo1	5U ExoIII/2U Exo1	2U ExoIII/2U Exo1
ExoIII	0.05 µl	0.025 µl	0.01 µl
ExoI	0.1 µl	0.1 µl	0.1 µl
T/E	0.85	0.875	0.89

One µl of the above reaction mixture was added to the appropriate tubes. The samples then incubated at 37°C for 14 minutes followed by 2 minutes @ 95°C and finally 1 minute @ 37°C

PCR was performed with universal primers by adding the following:

10x PCR buffer	2.5 µl (Promega)
water	19.0 µl
Universal P1	1.0 µl
Universal P2	1.0 µl

dNTPs	1.2 µl
Taq	0.3 µl

95°C 30s, 48°C 45s, 72°C 45s (25 cycles).

Detection of methylated, unmethylated and a mixture of 50% methylated + 50% unmethylated DNA

As can be seen from Figure 2 the system can easily differentiate methylated sequences from unmethylated and can also detect the presence of mixed populations.

One ng of either a methylated oligonucleotide, unmethylated oligo or 50%/50% mixture of methylated and unmethylated oligo was used as a target. Ten ng of each clamp was used in the reaction, together with 1 unit of Stoffel DNA polymerase. Ampligase enzyme was used at 5 units. The arrow indicates the correct size of the PCR products

In detail, the following was added to a 0.5 µl thin-walled PCR tube:

10x Ampligase buffer	1.0 µl
Clamp#1 (10 ng/µl)	1.0 µl
Clamp#2 (10 ng/µl)	1.0 µl
Stoffel Polymerase	0.1 µl (1 Unit)
Ampligase	1.0 µl
dG/dA/dT/dC (1 mM)	1.0 µl (dG/dA/dT/dC are added to separate tubes)
water	3.9 µl

Reactions were incubated as follows:

20°C 4 min, 95°C 2 min, 50°C 15 min, 55°C 5 min (with a 0.05°C/sec ramp time), 50°C 1 min.

Exonuclease treatment:

5U ExoIII/2U Exo1

ExoIII	0.025 µl
Exol	0.1 µl
T/E	0.875

One µl of the above reaction mixture was added to the appropriate tubes. The samples then incubated at 37°C for 14 minutes followed by 2 minutes @ 95°C and finally 1 minute @ 37°C

PCR was performed with universal primers by adding the following:

10x PCR buffer	2.5 µl (Promega)
water	19.0 µl
Universal P1	1.0 µl
Universal P2	1.0 µl
dNTPs	1.2 µl
Taq	0.3 µl

95°C 30s, 48°C 45s, 72°C 45s (25 cycles).

Effect of polymerase on the clamp reaction

Experiment carried out using methylated DNA sequences.

Figure 3 shows the effect of different polymerase enzymes on the performance of the reaction. It should be noted that all polymerase enzymes tested produced positive signals when the concentration of the target was further increased. However, not all enzymes could be used at low DNA concentrations without further optimisation.

The results showed that the enzymes that produced the greatest sensitivity were Klenow DNA polymerase and Taq Polymerase. These two enzymes were used subsequently in further experiments.

One ng of each clamp was used in the reaction, together with 1 units of DNA polymerase. The arrow indicates the correct size of the PCR products.

In detail, the following was added to a 0.5 µl thin-walled PCR tube:

10x Ampligase buffer	1.0 µl
Clamp#1 (1 ng/µl)	1.0 µl
Clamp#2 (1 ng/µl)	1.0 µl
Polymerase	1U
Ampligase	5U
dG/dA/dT (1 mM)	1.0 µl (dG/dA/dT are added to separate tubes)

water 3.9 µl

Template DNA was added at the appropriate concentrations indicated above.

Reactions were incubated as follows:

20°C 4 min, 95°C 2 min, 50°C 15 min, 55°C 5 min (with a 0.05°C/sec ramp time), 50°C 1 min.

Exonuclease treatment:

	5U ExoIII/2U Exo1
ExoIII	0.025 µl
ExoI	0.1 µl
T/E	0.875

One µl of the above reaction mixture was added to the appropriate tubes. The samples then incubated at 37°C for 14 minutes followed by 2 minutes @ 95°C and finally 1 minute @ 37°C

0.5 µl of uracil DNA glycosylase was added to each reaction tube and the samples incubated as follows; 50°C 15 min, 95°C 5 min, 50°C 1 min.

PCR was performed with universal primers by adding the following:

10x PCR buffer	2.5 µl (Promega)
water	19.0 µl
Universal P1	1.0 µl
Universal P2	1.0 µl
dNTPs	1.2 µl
Taq	0.3 µl

95°C 30s, 48°C 45s, 72°C 45s (25 cycles).

Effect of Polymerase/Ligase concentration

Experiment carried out using methylated DNA

Figure 4 shows that high concentrations of both enzymes can produce false positive signals such as those seen with 5 Units of both ligase and polymerase where bands are present in both the G and T tracks. Whereas using 2 Units ligase and 1 Unit polymerase a band is only present in the correct G track. Careful titration of the enzyme

combinations is therefore essential to produce the most sensitive and specific enzyme combinations. As can be seen 2U of ligase and 1U of Klenow gave the most consistent and reliable amplification even at low DNA concentrations.

One ng of each clamp was used in the reaction, together with the Ampligase and polymerase enzyme were used as above. The arrow indicates the correct size of the PCR products.

In detail, the following was added to a 0.5 µl thin-walled PCR tube:

10x Ampligase buffer	1.0 µl
Clamp#1 (1 ng/µl)	1.0 µl
Clamp#2 (1 ng/µl)	1.0 µl
Klenow	as above
Ampligase	as above
dG/dT (1 mM)	1.0 µl (dG/dT are added to separate tubes)
water	3.9 µl

Template DNA was added at the appropriate concentrations indicated above.

Reactions were incubated as follows:

20°C 4 min, 95°C 2 min, 50°C 15 min, 55°C 5 min (with a 0.05°C/sec ramp time), 50°C 1 min.

Exonuclease treatment:

	5U ExoIII/2U Exo1
ExoIII	0.025 µl
ExoI	0.1 µl
T/E	0.875

One µl of the above reaction mixture was added to the appropriate tubes. The samples then incubated at 37°C for 14 minutes followed by 2 minutes @ 95°C and finally 1 minute @ 37°C

0.5µl of uracil DNA glycosylase was added to each reaction tube and the samples incubated as follows; 50°C 15 min, 95°C 5 min, 50°C 1 min.

PCR was performed with universal primers by adding the following:

10x PCR buffer	2.5 µl (Promega)

water	19.0 µl
Universal P1	1.0 µl
Universal P2	1.0 µl
dNTPs	1.2 µl
Taq	0.3 µl

95°C 30s, 48°C 45s, 72°C 45s (25 cycles).

Sensitivity of the clamp reaction using clamp primers directed against methylated DNA sequences

As can be seen from Figure 5, the assay had very high specificity and sensitivity at all DNA concentrations tested.

One ng of each clamp was used in the reaction, together with the Ampligase enzyme was used at 2.5 and 1 units. The arrow indicates the correct size of the PCR products.

In detail, the following was added to a 0.5 µl thin-walled PCR tube:

10x Ampligase buffer	1.0 µl
Clamp#1 (1 ng/µl)	1.0 µl
Clamp#2 (1 ng/µl)	1.0 µl
Clamp#3 (1 ng/µl)	1.0 µl
Ampligase	0.5 or 0.2 µl
water	4.5 or 4.8 µl

One µl of template DNA was added at the appropriate concentrations indicated above.

Reactions were incubated as follows:

20°C 4 min, 95°C 2 min, 50°C 15 min, 55°C 5 min (with a 0.05°C/sec ramp time), 50°C 1 min.

Exonuclease treatment:

	5U ExoIII/2U Exo1
ExoIII	0.025 µl
ExoI	0.1 µl

T/E	0.875

One µl of the above reaction mixture was added to the appropriate tubes. The samples then incubated at 37°C for 14 minutes followed by 2 minutes @ 95°C and finally 1 minute @ 37°C

0.5µl of uracil DNA glycosylase was added to each reaction tube and the samples incubated as follows; 50°C 15 min, 95°C 5 min, 50°C 1 min.

PCR was performed with universal primers by adding the following:

10x PCR buffer	2.5 µl (Promega)
water	19.0 µl
Universal P1	1.0 µl
Universal P2	1.0 µl
dNTPs	1.2 µl
Taq	0.3 µl

95°C 30s, 48°C 45s, 72°C 45s (25 cycles).

Clamp reaction on genomic DNA (known to be unmethylated at the site of interrogation)

Ten pg or 1 pg of each clamp was used in the reaction, together with the ampligase enzyme at 2.5 and 1 units. The arrow indicates the correct size of the PCR products (Figure 6).

In detail, the following was added to a 0.5 µl thin-walled PCR tube:

10x Ampligase buffer	1.0 µl
Clamp#1	1.0 µl
Clamp#2	1.0 µl
Ampligase	0.5 µl
Polymerase	0.1 µl
dG/dA/dT/dC (1 mM)	1.0 µl (dG/dA/dT/dC are added to separate tubes)
water	4.4 µl

One µl of bisulphite template DNA was added.

Reactions were incubated as follows:

20°C 4 min, 95°C 2 min, 50°C 15 min, 55°C 5 min (with a 0.05°C/sec ramp time), 50°C 1 min.

Exonuclease treatment:

	5U ExoIII/2U Exo1
ExoIII	0.025 µl
ExoI	0.1 µl
T/E	0.875

One µl of the above reaction mixture was added to the appropriate tubes. The samples then incubated at 37°C for 14 minutes followed by 2 minutes @ 95°C and finally 1 minute @ 37°C

0.5 µl of uracil DNA glycosylase was added to each reaction tube and the samples incubated as follows; 50°C 15 min, 95°C 5 min, 50°C 1 min.

PCR was performed with universal primers by adding the following:

10x PCR buffer	2.5 µl (Promega)
water	19.0 µl
Universal P1	1.0 µl
Universal P2	1.0 µl
dNTPs	1.2 µl
Taq	0.3 µl

95°C 30s, 48°C 45s, 72°C 45s (40 cycles).

Clamp reaction - non-circular

Figure 7 shows the results of a ligase fill reaction using 2 single arms of two clamps.

Five pg of a uracil containing oligonucleotide (unmethylated) was used as a target. One fmol of each oligo "arm" was used in the reaction, together with 0.05 units of Taq polymerase. Ampligase enzyme was diluted from 1 unit to 0.03 unit. The arrow indicates the correct size of the PCR products.

In detail, the following was added to a 0.5 µl thin-walled PCR tube:

10x Ampligase buffer 0.9 µl

Oligo target (5 pg/µl)	1.0 µl
Clamp 1 (1 fmol/µl)	1.0 µl
Clamp 2 (1 fmol/µl)	1.0 µl
Taq Polymerase	0.1 µl (of a 0.5U/µl stock diluted in Ampligase buffer)
dA/dC/dG (1 mM)	1.0 µl (dA/dC/dG are added to separate tubes)
water	4.0 µl

Reactions were incubated as follows:

20°C 4 min, 95°C 5 min, 60°C 25 min, 65°C 5 min (with a 0.05°C/sec ramp time), 50°C 1 min.

One µl of UDG was added at this point to degrade the template.

50°C 15 min, 95°C 5 min, 50°C 1 min.

PCR was performed with universal primers by adding the following:

10x PCR buffer	2.5 µl (Promega)
water	19.0 µl
Universal P1	1.0 µl
Universal P2	1.0 µl
dNTPs	1.2 µl
Taq	0.3 µl

95°C 30s, 48°C 45s, 72°C 45s (28 cycles).

First aspect examples

Figure 8 show the method according to the first aspect of the present invention. In Example #1 and #2, the two adjacent clamp sequences were then ligated together using a DNA ligase.

Second aspect examples

Figure 9 shows the method according to the second aspect of the present invention. In Example #3, a polymerase enzyme will bring the correct base (G) to the site and then a ligase enzyme will ligate the base into the DNA sequence to form a linear probe. In Example #4, a polymerase enzyme will bring the correct base (A) to the site

and then a ligase enzyme will ligate the base into the DNA sequence to form a linear probe.

Third aspect examples

In Example #5, four independent hybridisation events need to occur to bring the three clamp sequences positioned correctly on the bisulphite treated genomic DNA (Figure 10). One arm of the clamps now form a semi-circular probe, while the other two clamps hybridise to the top and bottom strands of the DNA adjacent to the semi-circular probe. A ligase enzyme will ligate the base into the DNA sequence to form a fully linearised probe.

In Example #6, four independent hybridisation events need to occur to bring the three clamp sequences positioned correctly on the bisulphite treated genomic DNA (Figure 10). One arm of the clamps now form a semi-circular probe, while the other two clamps hybridise to the top and bottom strands of the DNA adjacent to the semi-circular probe. A ligase enzyme will ligate the base into the DNA sequence to form a fully linearised probe.

Fourth aspect examples

In Example #7, four independent hybridisation events need to occur to bring the three clamp sequences positioned correctly on the bisulphite treated genomic DNA (Figure 11). One arm of the clamp now forms a semi-circle probe, while the other two clamps hybridise to the top and bottom strands of the DNA adjacent to the semi-circular probe. A polymerase enzyme will bring the correct base (G) to the site and then a ligase enzyme will ligate the base into the DNA sequence to form a fully linearised probe.

In Example #8, four independent hybridisation events need to occur to bring the three clamp sequences positioned correctly on the bisulphite treated genomic DNA (Figure 11). One arm of the clamp now form a semi-circle probe, while the other two clamps hybridise to the top and bottom strands of the DNA adjacent to the semi-circular probe. A polymerase enzyme will bring the correct base (A) to the site and then a ligase enzyme will ligate the base into the DNA sequence to form a fully linearised probe.

Fifth aspect examples

In Example #9, four independent hybridisation events need to occur to bring the two clamp sequences positioned correctly on the bisulphite treated genomic DNA. Each arm of the clamps now form a semi-circle probe (Figure 12). A ligase enzyme will ligate the bases into the DNA sequence to form a fully circularised probe.

In Example #10, four independent hybridisation events need to occur to bring the two clamp sequences positioned correctly on the bisulphite treated genomic DNA. Each arm of the clamp probes now form a semi-circle (Figure 12). A ligase enzyme will ligate the base into the DNA sequence to form a fully circularised probe.

Sixth aspect examples

In Example #11, four independent hybridisation events need to occur to bring the two clamp sequences positioned correctly on the bisulphite treated genomic DNA. Each arm of the clamps now forms a semi-circle probe (Figure 13). A polymerase enzyme will bring the correct base (G) to the site and then a ligase enzyme will ligate the two semi-circular probes to from a fully circularised probe.

In Example #12, four independent hybridisation events need to occur to bring the two clamp sequences positioned correctly on the bisulphite treated genomic DNA (Figure 13). Each arm of the clamps now form a semi-circle probes. A polymerase enzyme will bring the correct base (A) to the site and then a ligase enzyme ligate the two semi-circular probes to from a fully circularised probe.

Figure 14 shows a schematic example of a clamp reaction forming a circular probe with a uracil cleavage site.

Array tag detection system

The bisulphite clamp is a new method for genome wide detection of methylated cytosine bases. Unlike all previous methods for the detection of methylated cytosines, the clamp method does not depend on the amplification of the bisulphite treated nucleic acid. The method is unique in that it can use a ligation reaction to determine the presence or absence of a methylated cytosine within a CpG doublet (see Figures 15 and 16). This means that individual CpG sites can be analysed throughout the genome.

After the clamps have been ligated and amplified they can then directed to specific sites on a micro-array devise predetermined by the internal Zip-code contained

within each bisulphite clamp primer set. Figure 16 shows an example of such a detection system.

One of the major advantages of this approach over conventional micro-array analysis is that the hybridisation conditions can be controlled in a precise manner. The zip code primers can be tailor made so that they all share the same melting temperatures and have similar sequence content. This means that the hybridisation kinetics will be very similar for each probe, whereas in conventional micro-array the hybridisations are dependant on the DNA sequence of the genomic nucleic acid and cannot be controlled in the same manner.

SUMMARY

There are a number of unique advantages to the bisulphite clamp method according to the present invention for the genome wide detection of methylation states compared to existing technologies. One major advantage is the exquisite specificity generated during the reaction. Conventional bisulphite analysis (Clark et al, 1994) relies on two rounds of nested PCR to generate the specificity required in a bisulphite converted genome. This is due to the fact that after bisulphite treatment the genome is essentially composed of only three bases namely A G and T since all the unmethylated cytosines have been converted to U. Four nested primers are then designed to one of the bisulphite treated strands of DNA then two rounds of PCR are performed to generate the necessary specificity and sensitivity. After bisulphite treatment, the two strands of DNA are no longer complementary. The targeting of both strands after conversion has to date not been utilised. The bisulphite clamp methodology according to the present invention can target both strands of the modified or converted DNA (see Figure 1), thus four independent hybridization reactions need to occur before the reaction will proceed. This in effect gives nested PCR specificity in one round reducing the chances of PCR contamination and greatly simplifying the procedure. In addition, non-specific amplification often occurs especially when the amount of target DNA is low. This is due to the nature of the PCR amplification reaction. The bisulphite clamp method, however, overcomes this by the use of a dual specificity system. The first layer of specificity is generated by the polymerase filling in the correct base. Even if the polymerase makes an error, the second round of fidelity is achieved by the use of a ligase enzyme which means that any incorrect base will not be correctly ligated and the reaction will come to a halt.

Another feature of the bisulphite clamp method can be the incorporation of a uracil site into one arm of one primer. Thus, when the correct hybridisation and ligation occurs a circle is formed. The reaction mix can then be treated with Uracil N Glycosylase (UNG). This enzyme will then degrade all the uracil-containing bisulphite treated DNA and at the same time linearise the circle. Thus, subsequent amplification now occurs in a pure population of target molecules greatly increasing the efficiency of the reaction unlike conventional PCR which occurs in a sea of non target genomic DNA. Alternatively, a unique restriction site can be added to one of the clamp primers so that after UNG digestion, the circles remain undigested but the bisulphite treated genomic DNA has been degraded. The circular DNA can then be replicated by isothermal Rolling Circle Amplification (RCA) (Lizardi et al. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. 1998 Nat. Genetics (19); 225-232).

Further, the bisulphite clamp method allows the multiplexing of a large number of probes in the same reaction tube. Thus, large numbers of genomic regions can be interrogated simultaneously in the one reaction tube. This is due to the fact that the clamps can contain common PCR primer sites, thus only two primers are required to amplify the target circularized molecules unlike multiplex PCR where individual primer sets need to be designed to each loci of interest.

In addition, the method also eliminates the problem of PCR bias which can often occur in a bisulphite reaction. This is a result of the fact that the bisulphite clamp does not depend on the amplification of a genomic region spanning two primers sites. It has been shown that amplification of certain sequences of bisulphite treated DNA can lead to an under-estimation of the true methylation status (Warneke et al). This phenomenon is a consequence of the differing sequence content of a methylated DNA strand compared to an unmethylated strand, resulting in the preferential copying of the unmethylated strand over the methylated strand. As the clamp only depends on a hybridization reaction, PCR bias will not occur.

Moreover, the detection of methylation changes in bisulphite treated DNA by micro-arrays has relied on sequence specific hybridisation of regions of amplified DNA contained between two unique primer sites. As stated previously, the sequence content of a methylated DNA sequence is very different from an unmethylated sequence. This has the effect that the hybridisation kinetics of an unmethylated sequence will be very different to that of a methylated sequence. Thus, in practical terms, it is very difficult to determine hybridisation conditions that will favour both the binding of the unmethylated sequence and the methylated sequence to an array simultaneously to the same array.

The bisulphite clamp procedure can overcome this problem by optionally incorporating zip-code sequence into the primers that subsequently direct the sequence to a defined place on an array. The zip-code sequence is not designed to correspond with a piece of bisulphite treated DNA but a randomly defined DNA sequence that is not present in the human genome. This zip-code sequence can be "shuffled" so that each zip code has the same melting temperature and sequence composition. Thus, the hybridisation conditions in the bisulphite clamp can be tailor made so that the kinetics of hybridisation of each probe will be identical. Thus the final read out will be a true reflection of the methylation status of the sample.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Claims:

1. A method for determining the methylation status of a potential methylation site in genomic nucleic acid comprising:

treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template containing a potential methylation site;

providing a first clamp containing a first capture sequence complementary to a region flanking one side of the potential methylation site in the modified nucleic acid template;

providing a second clamp containing a second capture sequence complementary to a region flanking the other side of the potential methylation site in the modified nucleic acid template;

allowing the first clamp and the second clamp to hybridise to the modified nucleic acid template;

ligating the hybridised first and second clamps to form a probe spanning the potential methylation site in the modified nucleic acid template;

digesting the modified nucleic acid template to obtain the probe; and

detecting the probe and determining the methylation status of the potential methylation site in the modified genomic nucleic acid.

2. The method according to claim 1 wherein the treating step forms a modified nucleic acid template containing two complementary strands of nucleic acid containing a potential methylation site; and

the first clamp contains a first capture sequence and a second capture sequence, first capture sequence being complementary to a region flanking one side of the potential methylation site in the first strand of the modified nucleic acid template, and the second capture sequence being complementary to a region flanking one side of the potential methylation site in the second strand of the modified nucleic acid template.

3. The method according to claim 1 or 2 wherein the first clamp and the second clamp hybridise to the modified nucleic acid template such that there is no complementary bases at the potential methylation site and causing at least one base to be inserted between the first clamp and the second clamp prior to ligating the clamps to form a probe spanning the potential methylation site in the modified nucleic acid template.

4. The method according to any one of claims 1 to 3 wherein the potential methylation site is cytosine (C) flanked 5' by a guanine (G).

5. The method according to any one of claims 1 to 4 wherein the modifying agent is selected from the group consisting of bisulfite, acetate, and citrate.

6. The method according to claim 5 wherein the modifying agent is sodium bisulfite .

7. The method according to any one of claims 1 to 6 wherein at least one of the first or second clamps contains a universal primer which allows amplification of the probe after the ligation step.

8. The method according to any one of claims 1 to 7 wherein at least one of the first or second clamps contains a capture site which allows capture of the probe.

9. The method according to any one of claims 1 to 8 wherein at least one of the first or second clamps contains a cleavable site.

10. The method according to claim 9 wherein the cleavable site is nuclease restriction site.

11. The method according to claim 10 wherein the cleavable site is a uracil base.

12. The method according to any one of claims 1, 2 and 4 to 11 wherein the complementary sequences of the first or second clamps span the potential methylation site.

13. The method according to any one of claims 1 to 12 wherein the clamps are ligated using a suitable enzyme capable of ligating single stranded nucleic acid.

14. The method according to any one of claims 1, 2 and 4 to 13 wherein the clamps are ligated at or near the first and second capture sequences.

15. The method according to any one of claims 3 to 13 wherein one or more bases are inserted between the clamps using a nucleic acid polymerase.

16. The method according to claim 15 wherein base insertion reaction is carried out in the presence of a single nucleotide type and each nucleotide type is reacted separately such that the insertion of a particular base will be indicative of the identity of complementary base on the modified nucleic acid template.

17. The method according to any one of claims 1 to 16 wherein the modified nucleic acid template is digested by an enzyme.

18. The method according to claim 17 wherein the enzyme is Uracil N DNA Glycosylase.

19. The method according to any one of claims 1 to 18 wherein the probe is detected and the methylation status of the potential methylation site in the modified genomic nucleic acid determined with a detection system which can recognise incorporation of

the missing base(s) between the first and second clamps, or detect specific amplification of the probe.

20. The method according to claim 19 wherein the detection system is selected from the group consisting of hybridization by array, Southern blot type detection, Real-Time PCR quantitation, fluorescent beads, enzyme conjugates, radioactive beads, ligase chain reaction, and isothermal DNA amplification technologies including strand displacement amplification or rolling circle amplification.

21. The method according to any one of claims 1 to 20 wherein the genomic nucleic acid is DNA.

22. A method for determining the methylation status of a potential methylation site on genomic nucleic acid comprising:

treating genomic nucleic acid with an agent which modifies cytosine bases but does not modify 5-methyl-cytosine bases under conditions to form a modified nucleic acid template containing two complementary strands of nucleic acid containing a potential methylation site;

providing a first clamp containing a first capture sequence and a second capture sequence, first capture sequence being complementary to a region flanking one side of the potential methylation site in the first strand of the modified nucleic acid template, and the second capture sequence being complementary to a region flanking one side of the potential methylation site in the second strand of the modified nucleic acid template;

providing a second clamp containing a third capture sequence and a fourth capture sequence, the third capture sequence being complementary to a region flanking the other side the potential methylation site in the first strand of the modified nucleic acid template, and the fourth capture sequence being complementary to a region flanking the other side of the potential methylation site in the second strand of the modified nucleic acid template;

allowing the first clamp and the second clamp to hybridise to the two complementary strands of the modified nucleic acid template;

ligating the hybridised clamps to form a circular probe spanning the potential methylation site in the complementary strands of the modified nucleic acid template;

digesting the modified nucleic acid template to obtain the circular probe; and

detecting the circular probe and determining the methylation status of the potential methylation site in the modified genomic nucleic acid.

23. The method according to claim 22 wherein the first clamp and the second clamp hybridise to the two complementary strands of the modified nucleic acid template such that there is no complementary bases at the potential methylation site of at least one of the complementary strands of the modified nucleic acid template, and causing at least one base to be inserted between the first clamp and the second clamp prior to ligating the clamps to form a probe spanning the potential methylation site in the modified nucleic acid template.

24. The method according to claim 22 or 23 wherein the potential methylation site is cytosine (C) flanked 5' by a guanine (G).

25. The method according to any one of claims 22 to 24 wherein the modifying agent is selected from the group consisting of bisulfite, acetate, and citrate.

26. The method according to claim 24 wherein the modifying agent is sodium bisulfite .

27. The method according to any one of claims 22 to 26 wherein at least one of the first or second clamps contains a universal primer which allows amplification of the probe after the ligation step.

28. The method according to any one of claims 22 to 27 wherein at least one of the first or second clamps contains a capture site which allows capture of the probe.

29. The method according to any one of claims 22 to 28 wherein at least one of the first or second clamps contains a cleavable site.

30. The method according to claim 29 wherein the cleavable site is nuclease restriction site.

31. The method according to claim 29 wherein the cleavable site is a uracil base.

32. The method according to any one of claims 22 and 24 to 31 wherein the complementary sequences of the first or second clamps span the potential methylation site.

33. The method according to any one of claims 22 to 32 wherein the clamps are ligated using a suitable enzyme capable of ligating single stranded nucleic acid.

34. The method according to any one of claims 22 and 24 to 33 wherein the clamps are ligated at or near the first and second capture sequences.

35. The method according to any one of claims 23 to 33 wherein one or more bases are inserted between the clamps using a DNA polymerase.

36. The method according to claim 35 wherein base insertion reaction is carried out in the presence of a single nucleotide type and each nucleotide type is reacted separately such that the insertion of a particular base will be indicative of the identity of complementary base on the modified nucleic acid template.

37. The method according to any one of claims 22 to 36 wherein the modified nucleic acid template is digested by an enzyme.

38. The method according to claim 37 wherein the enzyme is Uracil N DNA Glycosylase.

39. The method according to any one of claims 22 to 38 wherein the probe is detected and the methylation status of the potential methylation site in the modified genomic nucleic acid determined with a detection system which can recognise incorporation of the missing base(s) between the first and second clamps, or detect specific amplification of the probe.

40. The method according to claim 39 wherein the detection system is selected from the group consisting of hybridization by array, Southern blot type detection, Real-Time PCR quantitation, fluorescent beads, enzyme conjugates, radioactive beads, ligase chain reaction, and isothermal DNA amplification technologies including strand displacement amplification or rolling circle amplification.

41. The method according to any one of claims 22 to 40 wherein the genomic nucleic acid is DNA.

Figure 1A. 5 Units of Ligase enzyme

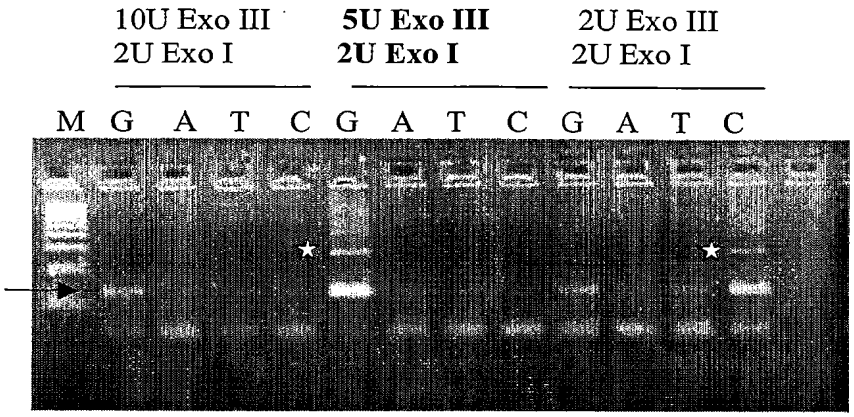


Figure 1B. 10 Units of Ligase enzyme

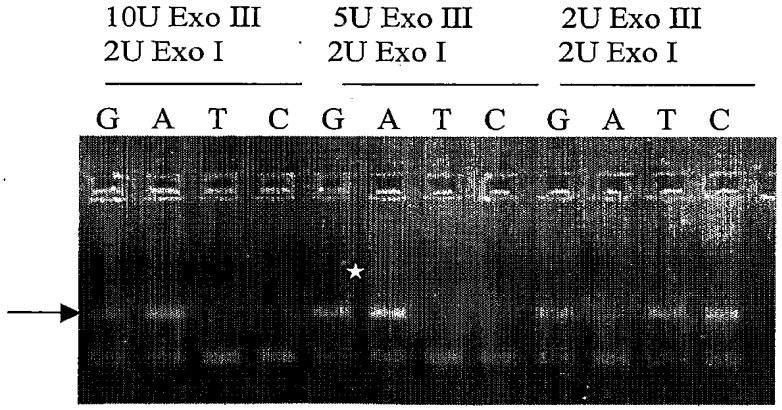


Figure 2

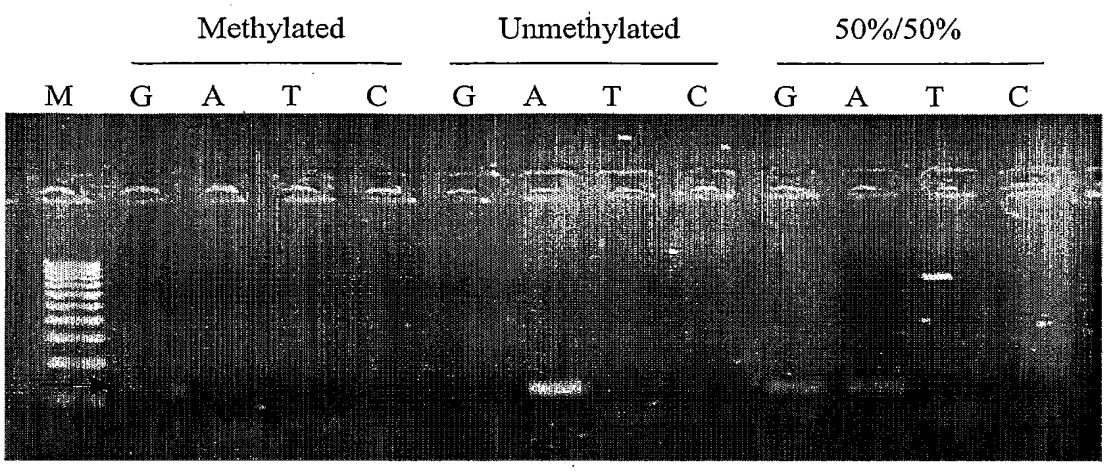
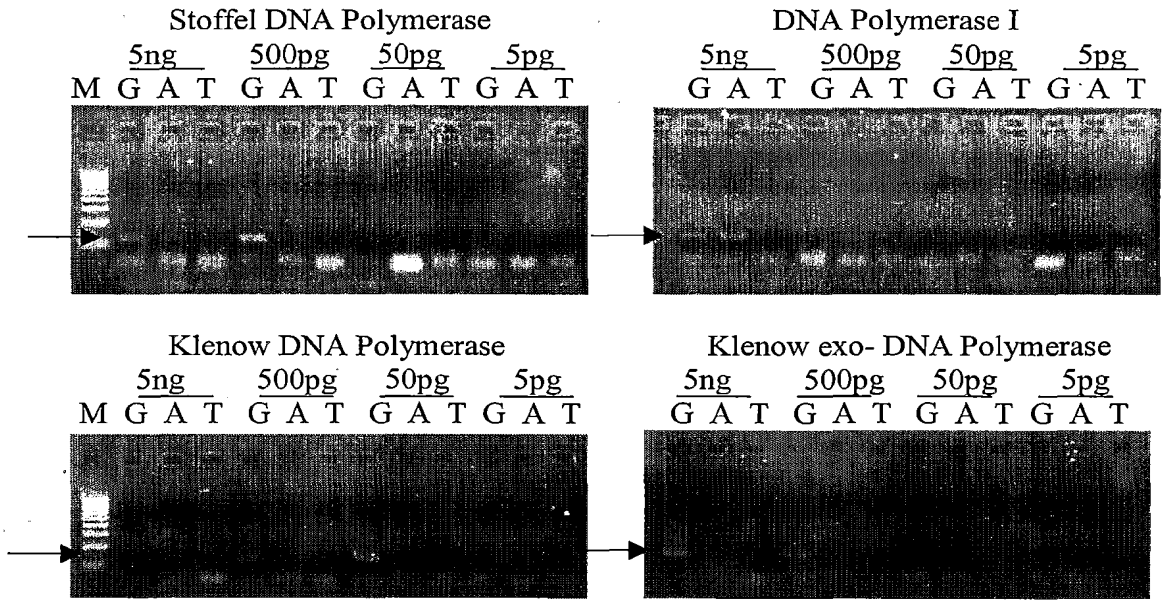


Figure 3

Panel A High concentration DNA



Panel B Low concentration DNA

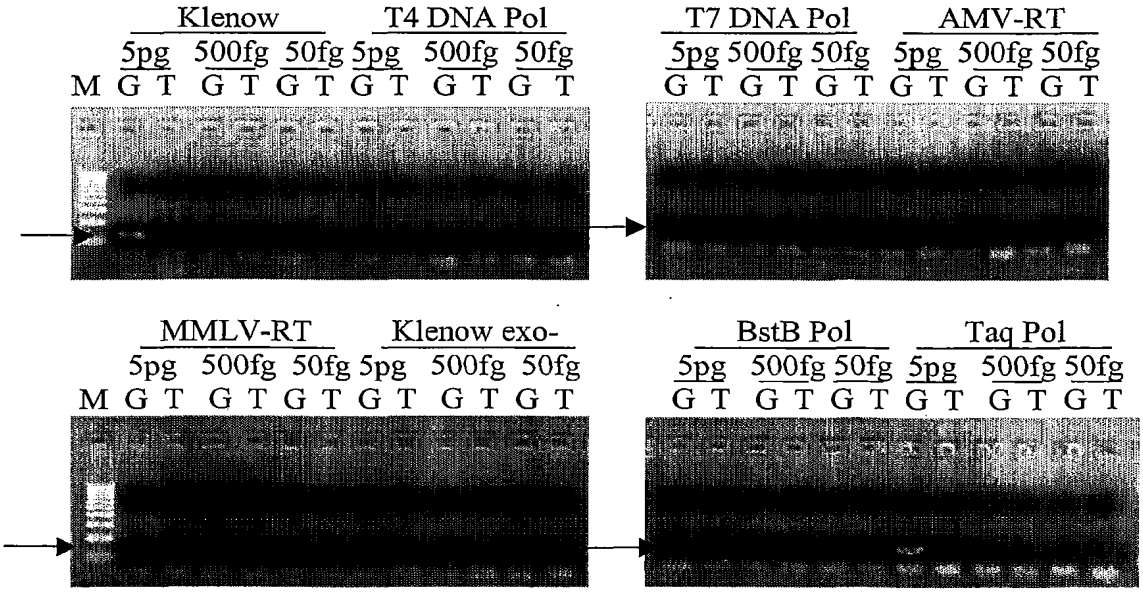


Figure 4 .

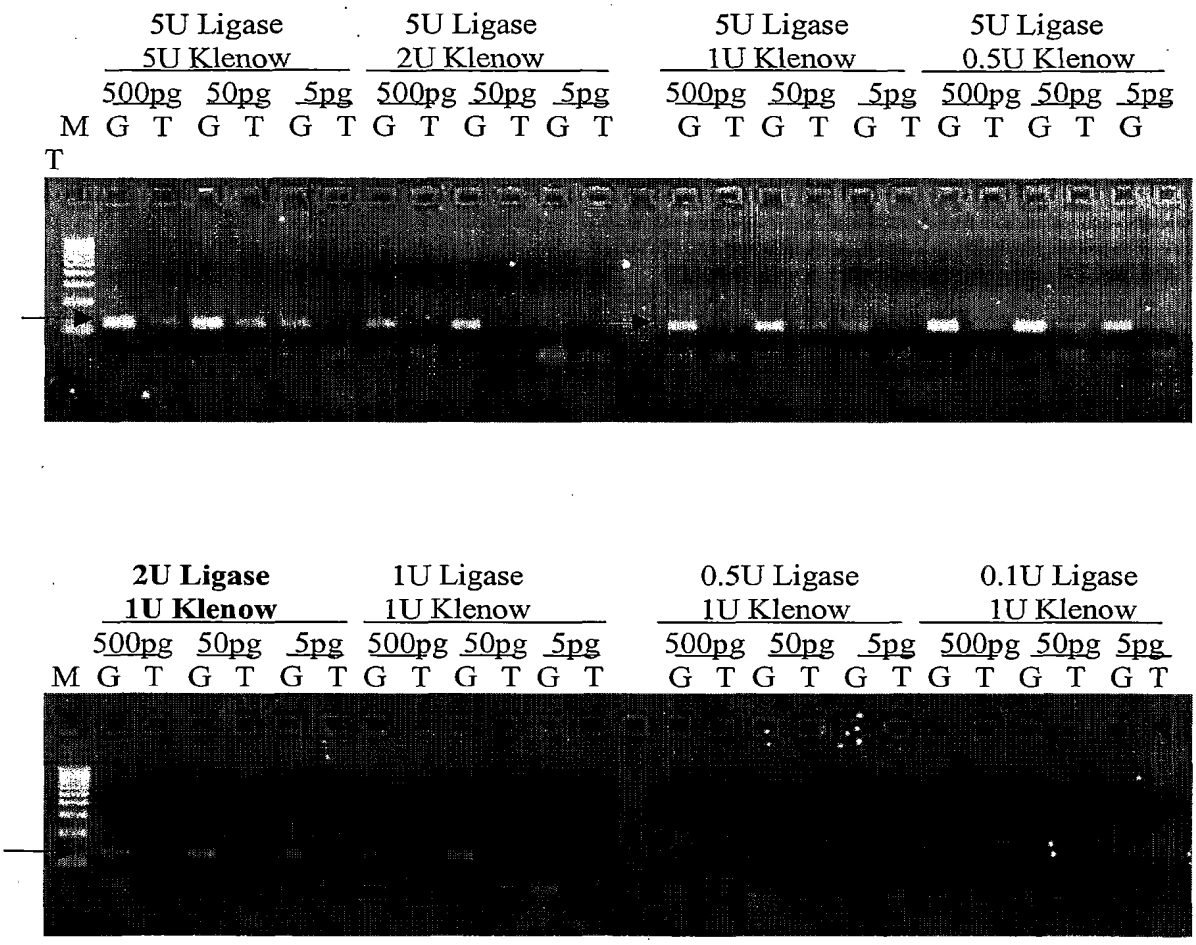
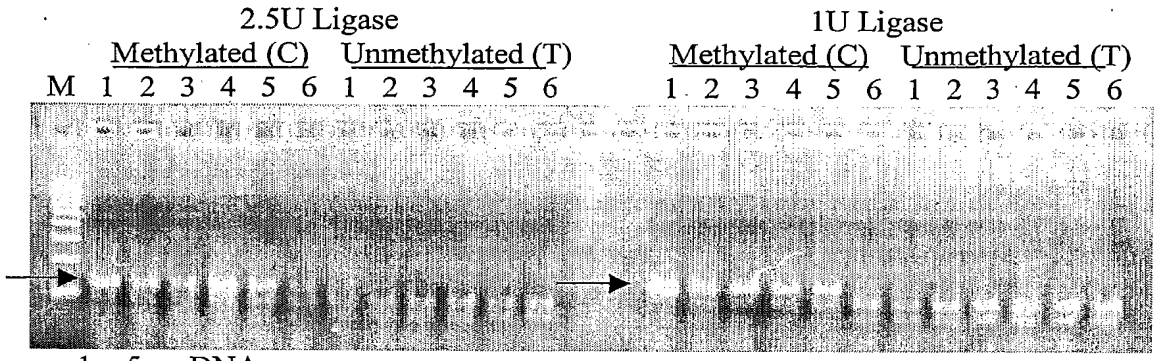

5U Ligase 5U Klenow
5U Ligase 2U Klenow
5U Ligase 1U Klenow
5U Ligase 0.5U Klenow
500pg 50pg 5pg 500pg 50pg 5pg 500pg 50pg 5pg 500pg 50pg 5pg
M G T G T G T G T G T G T G T G T G T G T G T G T G T G T G T
2U Ligase 1U Klenow
1U Ligase 1U Klenow
0.5U Ligase 1U Klenow
0.1U Ligase 1U Klenow
500pg 50pg 5pg 500pg 50pg 5pg 500pg 50pg 5pg 500pg 50pg 5pg
M G T G T G T G T G T G T G T G T G T G T G T G T G T G T G T

Figure 5



1. 5 ng DNA
2. 500 pg DNA
3. 50 pg DNA
4. 5 pg DNA
5. 500 fg DNA
6. 50 fg DNA

Figure 6

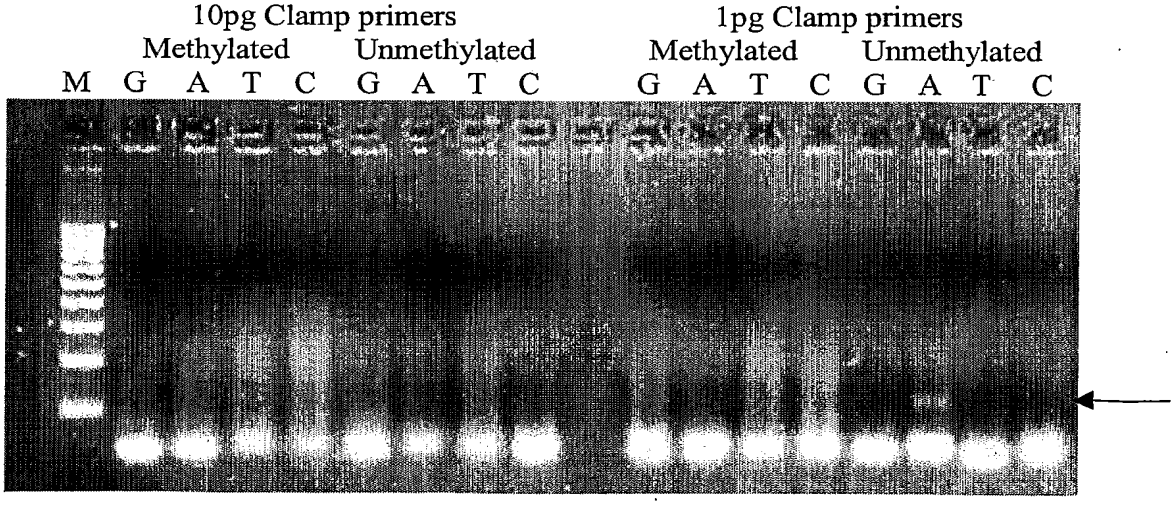


Figure 7

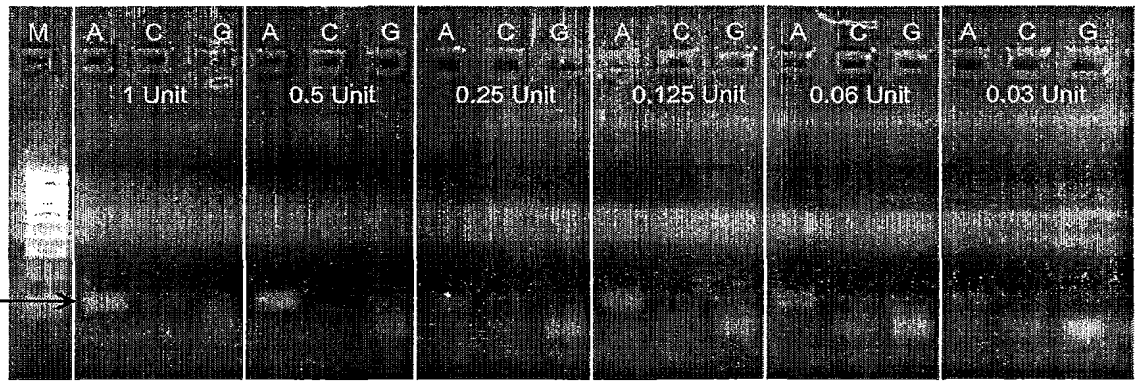

M
A C G
1 Unit
A C G
0.5 Unit
A C G
0.25 Unit
A C G
0.125 Unit
A C G
0.06 Unit
A C G
0.03 Unit

Figure 8
Example #1

Methylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG C GG AAA TTT YGT T-3'
Clamp#1 TCT TTT CTA TAA ACC G CC TTT AAA RCA A
Clamp#2

Example #2

Unmethylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG T GG AAA TTT YGT T-3'
Clamp#1 TCT TTT CTA TAA ACC A CC TTT AAA RCA A
Clamp#2

Figure 9

Example#3

Methylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG C GG AAA TTT YGT T-3'
Clamp#1 TCT TTT CTA TAA ACC CC TTT AAA RCA A Clamp#2

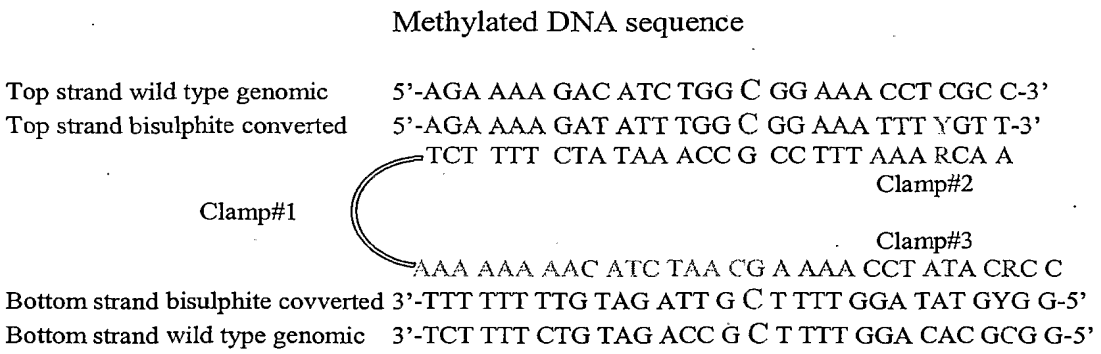
Example#4

Unmethylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG T GG AAA TTT YGT T-3'
Clamp#1 TCT TTT CTA TAA ACC CC TTT AAA RCA A Clamp#2

Figure 10

Example #5



Example #6

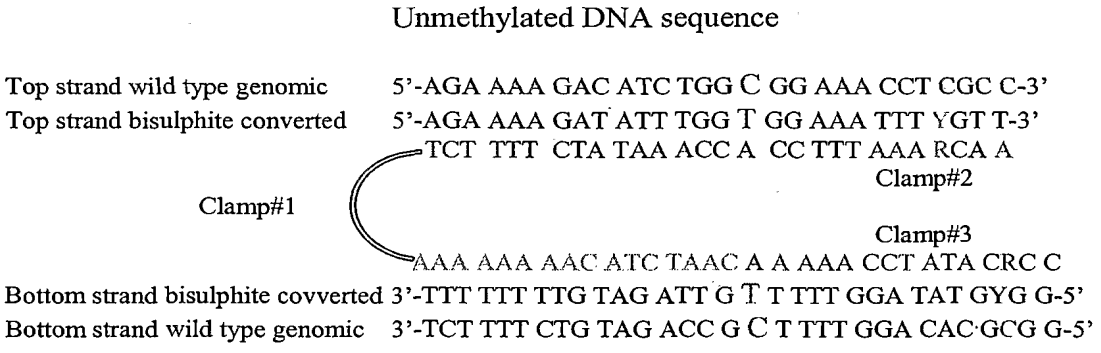
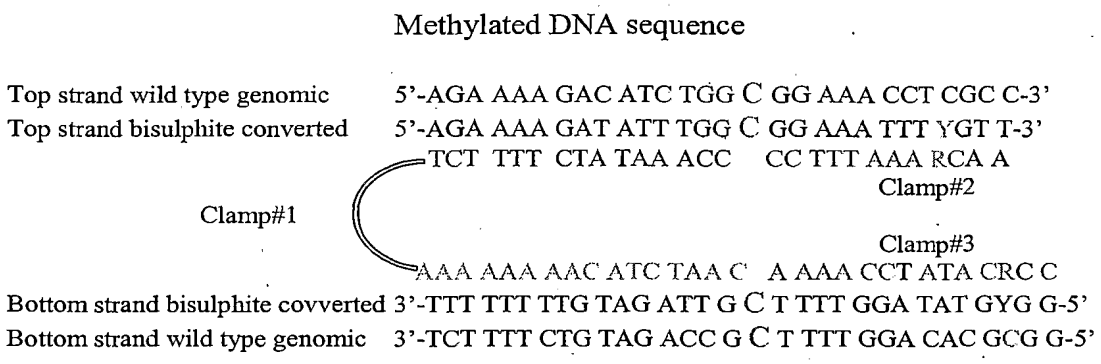


Figure 11

Example #7



Example #8



Figure 12

Example #9

Methylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG C GG AAA TTT YGT T-3'
TCT TTT CTA TAA ACC G CC TTT AAA RCA A
Clamp#1 Clamp#2
AAA AAA AAC ATC TAA CG A AAA CCT ATA CRC C
Bottom strand bisulphite covverted 3'-TTT TTT TTG TAG ATT G C T TTT GGA TAT GYG G-5'
Bottom strand wild type genomic 3'-TCT TTT CTG TAG ACC G C T TTT GGA CAC GCG G-5'

Example #10

Unmethylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG T GG AAA TTT YGT T-3'
TCT TTT CTA TAA ACC A CC TTT AAA RCA A
Clamp#1 Clamp#2
AAA AAA AAC ATC TAAC A A AAA CCT ATA CRC C
Bottom strand bisulphite covverted 3'-TTT TTT TTG TAG ATT G T T TTT GGA TAT GYG G-5'
Bottom strand wild type genomic 3'-TCT TTT CTG TAG ACC G C T TTT GGA CAC GCG G-5'

Figure 13

Example#11

Methylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG C GG AAA TTT YGT T-3'
TCT TTT CTA TAA ACC CC TTT AAA RCA A
Clamp#1 Clamp#2
AAA AAA AAC ATC TAA C A AAA CCT ATA CRC C
Bottom strand bisulphite covverted 3'-TTT TTT TTG TAG ATT G C T TTT GGA TAT GYG G-5'
Bottom strand wild type genomic 3'-TCT TTT CTG TAG ACC G C T TTT GGA CAC GCG G-5'

Example#12

Unmethylated DNA sequence

Top strand wild type genomic 5'-AGA AAA GAC ATC TGG C GG AAA CCT CGC C-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG T GG AAA TTT YGT T-3'
TCT TTT CTA TAA ACC CC TTT AAA RCA A
Clamp#1 Clamp#2
AAA AAA AAC ATC TAAC A AAA CCT ATA CRC C
Bottom strand bisulphite covverted 3'-TTT TTT TTG TAG ATT G T T TTT GGA TAT GYG G-5'
Bottom strand wild type genomic 3'-TCT TTT CTG TAG ACC G C T TTT GGA CAC GCG G-5'

Figure 14

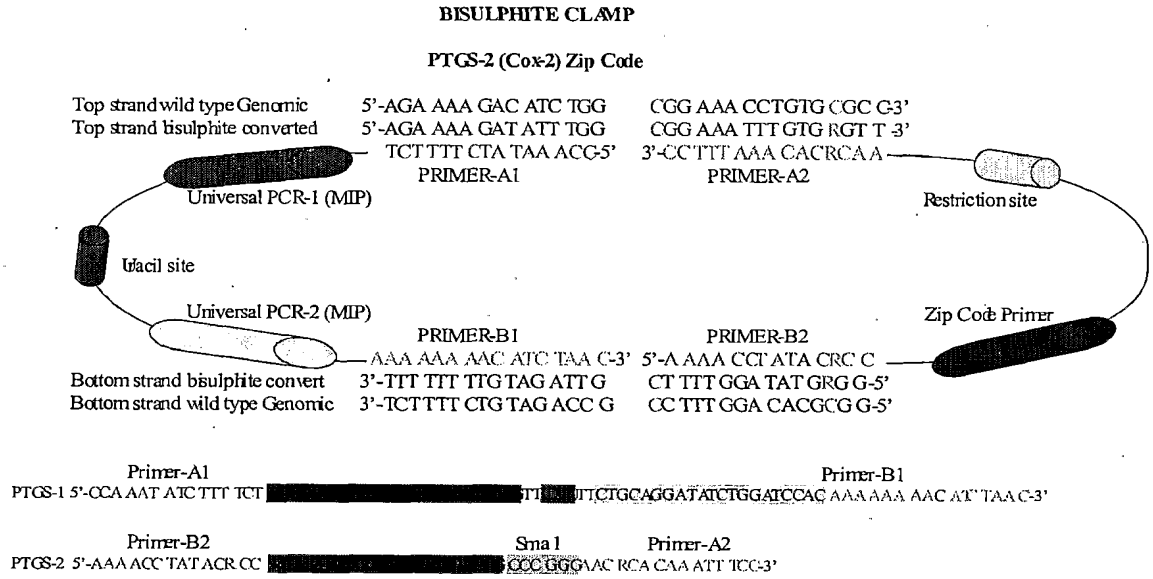

BISULPHITE CLAMP
PTGS-2 (Cox-2) Zip Code
Top strand wild type Genomic 5'-AGA AAA GAC ATC TGG CGG AAA CCTGTG CGC G-3'
Top strand bisulphite converted 5'-AGA AAA GAT ATT TGG CGG AAA TTT GTG RGT T-3'
TCT TTT CTA TAA ACG-5' 3'-CC TTT AAA CACRCA A
PRIMER-A1
PRIMER-A2
Universal PCR-1 (MIP)
Restriction site
Uracil site
Universal PCR-2 (MIP)
Zip Code Primer
PRIMER-B1
PRIMER-B2
AAA AAA AAC ATC TAA C-3' 5'-A AAA CCT ATA CRC C
Bottom strand bisulphite convert 3'-TTT TTT TTG TAG ATT G CT TTT GGA TAT GRG G-5'
Bottom strand wild type Genomic 3'-TCT TTT CTG TAG ACC G CC TTT GGA CACGCG G-5'
Primer-A1
PTGS-1 5'-CCA AAT ATC TTT TCT
TTCTGCAGGATATCTGGATCCAC AAA AAA AAC ATT TAA C-3'
Primer-B1
Primer-B2
Sma I
Primer-A2
PTGS-2 5'-AAA ACC TAT ACR CC
CCC GGG AAC RCA CAA ATT TCC-3'

Figure 15

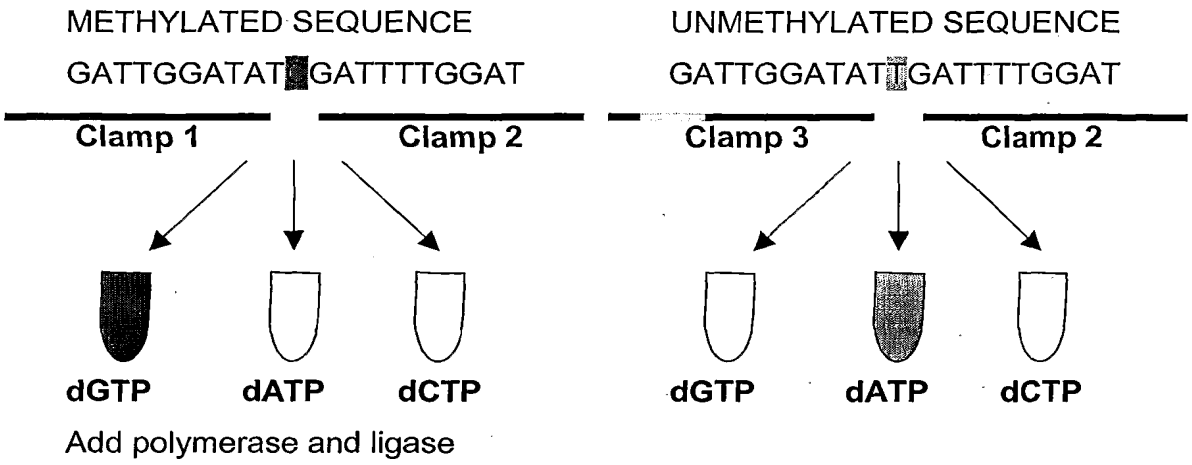
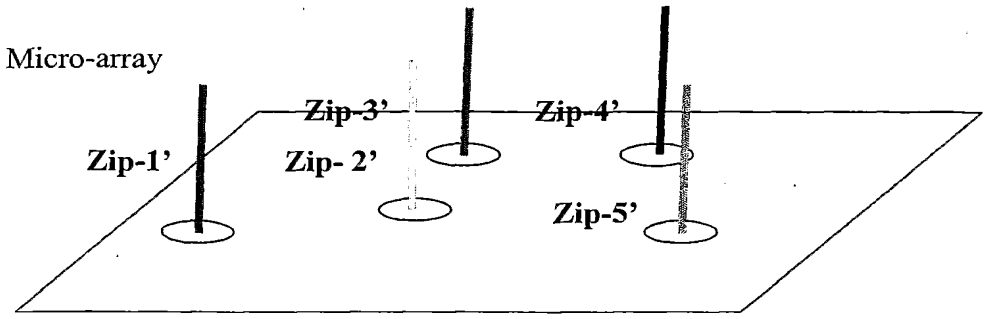


Figure 16

Zip-1	Clamp A
Zip-2	Clamp B
Zip-3	Clamp C
Zip-4	Clamp D
Zip-5	Clamp E



INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2004/001196

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: C12Q 1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPIDS Medline CA: DNA methylation/methylcytosine, bisulphite/bisulfite, probe, hybridise/anneal, flanking/adjacent, circularise, polymerase, ligase, exonuclease.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2003/0119025 A1 (Alexander OLEK and Kurt BERLIN) 26 June 2003	1, 3-11, 13-21
X	US 5 952 174 A (Theo NIKIFOROV et al) 14 September 1999.	1, 3-11, 13-21
X	WO 1997/41254 A1 (LANDERGREN, Ulf) 6 November 1997.	12
X	Nilsson et al. Padlock probes: circularising oligonucleotides for localised DNA detection. Science, 1994. 265: 2085-88.	12

☐ Further documents are listed in the continuation of Box C ☒ See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 September 2004	27 SEP 2004
Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	**Gillian Allen** Telephone No : (02) 6283 2266

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/AU2004/001196

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report	Patent Family Member
US 2003/0119025	AU 2001/42279
	CA 2401198
	EP 1261740
	WO 2001/62961
US 5952174	AU 1995/18407
	CA 2182517
	EP 0754240
	WO 1995/21271
WO 1997/41254	CA 2252817
	EP 0951565
	US 2001007742
	US 2002102592

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 December 2013 (19.12.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/188846 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C40B 30/04* (2006.01)
G01N 33/53 (2006.01)

(21) International Application Number:
PCT/US2013/046020

(22) International Filing Date:
14 June 2013 (14.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/660,427 15 June 2012 (15.06.2012) US

(72) Inventor; and
(71) Applicant : **STYLLI, Harry** [GB/US]; 2452 Paseo Dorado, La Jolla, CA 92037 (US).

(74) Agents: **HALEY, James, F.** et al.; Ropes & Gray LLP, 1211 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: METHODS OF DETECTING DISEASES OR CONDITIONS

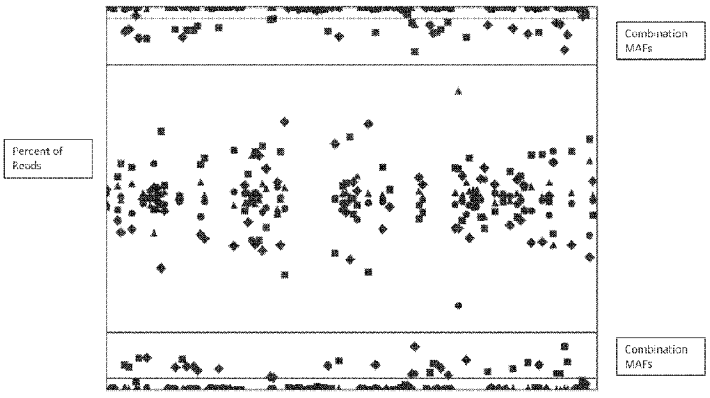


Figure 2

(57) Abstract: This invention provides methods of using a sample with multiple analytical components in the diagnosis, prognosis, or monitoring of diseases or conditions. The invention also provides methods of identifying markers of diseases or conditions.

WO 2013/188846 A1

METHODS OF DETECTING DISEASES OR CONDITIONS

Priority Information

[0001] This application claims priority from U.S. Provisional Application 61/660,427, filed June 15, 2012. The content and disclosure of that application is incorporated by reference herein in its entirety.

Field of the Invention

[0002] This invention relates generally to methods of using combinations of two or more different components selected from cell-free bodily fluids, phagocytic cells, circulating vesicles, and circulating diseased cells in the diagnosis, prognosis, or monitoring of a disease or condition. The invention also relates to methods of using the combinations to identify markers of diseases or conditions.

Background of the Invention

[0003] Leukocytes begin as pluripotent hematopoietic stem cells in the bone marrow and develop along either the myeloid lineage (monocytes, macrophages, neutrophils, eosinophils, and basophils) or the lymphoid lineage (T and B lymphocytes and natural killer cells). The major function of the myeloid lineage cells (e.g., neutrophils and macrophages) is the phagocytosis of infectious organisms, live unwanted damaged cells, senescent and dead cells (apoptotic and necrotic), as well as the clearing of cellular debris. Phagocytes from healthy animals do not replicate and are diploid, i.e., have a DNA content of 2n. On average, each cell contains <10 ng

DNA, <20 ng RNA, and <300 ng of protein. Non-phagocytic cells are also diploid and are not involved in the internalization of dead cells or infectious organisms and also have a DNA content of 2n.

[0004] The lifetime of various white blood cell subpopulations varies from a few days (e.g., neutrophils) to several months (e.g., macrophages). Like other cell types, leukocytes age and eventually die. During their aging process, human blood- and tissue-derived phagocytes (e.g., neutrophils) exhibit all the classic markers of programmed cell death (i.e., apoptosis), including caspase activation, pyknotic nuclei, and chromatin fragmentation. These cells also display a number of "eat-me" flags (e.g., phosphatidylserine, sugars) on the extracellular surfaces of their plasma membranes. Consequently, dying and dead cells and subcellular fragments thereof are cleared from tissues and blood by other phagocytic cells.

[0005] Early diagnosis of a disease often increases the likelihood of successful treatment or cure of such disease. Current diagnostic methods, however, focus on either using whole blood or separating the blood into different components, of which a single component is chosen for testing. Although this approach may enrich the signal being detected, it also results in the loss of potentially important information. Personalized diagnostic methods are needed that enable the diagnosis, especially the early diagnosis, of the presence of a disease or a condition in individuals who are not known to have the disease or who have recurrent disease.

[0006] One object of the present invention is to provide diagnostic methods that can facilitate the detection of a disease or condition-specific markers, e.g., nucleic acids, proteins, carbohydrates, and/or lipids and the like by using combinations of two or more different components selected from cell-free bodily fluids, phagocytic cells, circulating vesicles, and circulating diseased cells. Another object of this invention is to provide methods of identifying a disease or condition-specific markers and further use such markers alone or together with any known markers to diagnose diseases or conditions.

Summary of the Invention

[0007] Some embodiments of the invention are:

1. A method for diagnosing or aiding in the diagnosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject, a population of phagocytic cells isolated from the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject;

b) determining a second profile of at least one of the one or more markers from a control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject, a population of non-phagocytic cells isolated from the subject, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition; and

c) identifying a difference between the first and second profiles, wherein the difference is indicative of the presence of said disease or condition in the subject.

2. A method for assessing the risk of developing a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject, a population of phagocytic cells isolated from the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject;

b) determining a second profile of at least one of the one or more markers from a control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject, a population of non-phagocytic cells isolated from the subject, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition; and

c) identifying a difference between the first and second profiles, wherein the difference is indicative of the risk of developing said disease or condition in the subject.

3. A method for prognosing or aiding in the prognosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject, a population of phagocytic cells isolated from the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject;

b) determining a second profile of at least one of the one or more markers from a control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject, a population of non-phagocytic cells isolated from the subject, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition; and

c) identifying a difference between the first and second profiles, wherein the difference is indicative of the prognosis of said disease or condition in the subject.

4. A method for assessing the efficacy of a treatment for a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject before the treatment, a population of phagocytic cells isolated from the subject before the treatment, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before the treatment, a population of circulating vesicles isolated from the subject before the treatment, and a population of circulating diseased cells isolated from the subject before the treatment;

determining a second profile of at least one of the one or more markers from a first control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject before the treatment, a population of non-phagocytic cells isolated from the subject before the treatment, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, before the treatment;

identifying a difference between the first and second profiles;

b) determining a third profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject after the treatment, a population of phagocytic cells isolated from the subject after the treatment, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after the treatment, a population of circulating vesicles isolated from the subject after the treatment, and a population of circulating diseased cells isolated from the subject after the treatment;

determining a fourth profile of at least one of the one or more markers from a second control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject after the treatment, a population of non-phagocytic cells isolated from the subject after the treatment, and a population of control cells isolated

from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, after the treatment;

identifying a difference between the third and fourth profiles; and

c) identifying a difference between the difference identified in a) and the difference identified in b) wherein the identified difference in c) is indicative of the efficacy of the treatment for said disease or condition in the subject.

5. A method for monitoring the progression or regression of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject at a first time point, a population of phagocytic cells isolated from the subject at a first time point, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a first time point, a population of circulating vesicles isolated from the subject at a first time point, and a population of circulating diseased cells isolated from the subject at a first time point;

determining a second profile of at least one of the one or more markers from a first control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject at a first time point, a population of non-phagocytic cells isolated from the subject at a first time point, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, at a first time point;

identifying a difference between the first and second profiles;

b) determining a third profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject at a second time point, a population of phagocytic cells isolated from the subject at a

second time point, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a second time point, a population of circulating vesicles isolated from the subject at a second time point, and a population of circulating diseased cells isolated from the subject at a second time point;

determining a fourth profile of at least one of the one or more markers from a second control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject at a second time point, a population of non-phagocytic cells isolated from the subject at a second time point, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, at a second time point;

identifying a difference between the third and fourth profiles; and

c) identifying a difference between the difference identified in a) and the difference identified in b) wherein the identified difference in c) is indicative of the progression or regression of said disease or condition in the subject.

6. A method for identifying a compound capable of ameliorating or treating a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject before administering the compound to the subject, a population of phagocytic cells isolated from the subject before administering the compound to the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before administering the compound to the subject, a population of circulating vesicles isolated from the subject before administering the compound to the subject, and a population of circulating diseased cells isolated from the subject before administering the compound to the subject;

determining a second profile of at least one of the one or more markers from a first control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject before administering the compound to the subject, a population of non-phagocytic cells isolated from the subject before administering the compound to the subject, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, before administering the compound to the subject;

identifying a difference between the first and second profiles;

b) determining a third profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject after administering the compound to the subject, a population of phagocytic cells isolated from the subject after administering the compound to the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after administering the compound to the subject, a population of circulating vesicles isolated from the subject after administering the compound to the subject, and a population of circulating diseased cells isolated from the subject after administering the compound to the subject;

determining a fourth profile of at least one of the one or more markers from a second control comprising a component selected from the group consisting of: a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject after administering the compound to the subject, a population of non-phagocytic cells isolated from the subject after administering the compound to the subject, and a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, after administering the compound to the subject;

identifying a difference between the third and fourth profiles; and

c) identifying a difference between the difference identified in a) and the difference identified in b) wherein the identified difference in c) indicates that the compound is capable of ameliorating or treating said disease or condition in the subject.

7. A method for assessing the efficacy of a treatment for a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject before the treatment, a population of phagocytic cells isolated from the subject before the treatment, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before the treatment, a population of circulating vesicles isolated from the subject before the treatment, and a population of circulating diseased cells isolated from the subject before the treatment;

b) determining a second profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject after the treatment, a population of phagocytic cells isolated from the subject after the treatment, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after the treatment, a population of circulating vesicles isolated from the subject after the treatment, and a population of circulating diseased cells isolated from the subject; and

c) identifying a difference between the first profile and the second profile, wherein the identified difference is indicative of the efficacy of the treatment for said disease or condition in the subject.

8. A method for monitoring the progression or regression of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected

from the group consisting of: a cell-free bodily fluid isolated from the subject at a first time point, a population of phagocytic cells isolated from the subject at a first time point, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a first time point, a population of circulating vesicles isolated from the subject at a first time point, and a population of circulating diseased cells isolated from the subject at a first time point;

b) determining a second profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject at a second time point, a population of phagocytic cells isolated from the subject at a second time point, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a second time point, a population of circulating vesicles isolated from the subject at a second time point, and a population of circulating diseased cells isolated from the subject at a second time point; and

c) identifying a difference between the first profile and the second profile, wherein the identified difference is indicative of the progression or regression of said disease or condition in the subject.

9. A method for identifying a compound capable of ameliorating or treating a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject before administering the compound to the subject, a population of phagocytic cells isolated from the subject before administering the compound to the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before administering the compound to the subject, a population of circulating vesicles isolated from the subject before administering the compound to the subject, and a population of circulating diseased cells isolated from the subject before administering the compound to the subject;

b) determining a second profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject after administering the compound to the subject, a population of phagocytic cells isolated from the subject after administering the compound to the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after administering the compound to the subject, a population of circulating vesicles isolated from the subject after administering the compound to the subject, and a population of circulating diseased cells isolated from the subject after administering the compound to the subject; and

c) identifying a difference between the first profile and the second profile, wherein the identified difference indicates that the compound is capable of ameliorating or treating said disease or condition in the subject.

10. A method for diagnosing or aiding in the diagnosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject, an analyte isolated from a population of phagocytic cells isolated from the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, an analyte isolated from a population of circulating vesicles isolated from the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject;

b) determining a second profile of at least one of the one or more markers from a control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject, an analyte isolated from a population of non-phagocytic cells isolated from the subject, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition; and

c) identifying a difference between the first and second profiles, wherein the difference is indicative of the presence of said disease or condition in the subject.

11. A method for assessing the risk of developing a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject, an analyte isolated from a population of phagocytic cells isolated from the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, an analyte isolated from a population of circulating vesicles isolated from the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject;

b) determining a second profile of at least one of the one or more markers from a control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject, an analyte isolated from a population of non-phagocytic cells isolated from the subject, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition; and

c) identifying a difference between the first and second profiles, wherein the difference is indicative of the risk of developing said disease or condition in the subject.

12. A method for prognosing or aiding in the prognosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject, an analyte isolated from a population of phagocytic cells isolated from the

subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, an analyte isolated from a population of circulating vesicles isolated from the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject;

b) determining a second profile of at least one of the one or more markers from a control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject, an analyte isolated from a population of non-phagocytic cells isolated from the subject, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition; and

c) identifying a difference between the first and second profiles, wherein the difference is indicative of the prognosis of said disease or condition in the subject.

13. A method for assessing the efficacy of a treatment for a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject before the treatment, an analyte isolated from a population of phagocytic cells isolated from the subject before the treatment, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before the treatment, an analyte isolated from a population of circulating vesicles isolated from the subject before the treatment, and an analyte isolated from a population of circulating diseased cells isolated from the subject before the treatment;

determining a second profile of at least one of the one or more markers from a first control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n

(=2n phagocytic cells) isolated from the subject before the treatment, an analyte isolated from a population of non-phagocytic cells isolated from the subject before the treatment, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, before the treatment;

identifying a difference between the first and second profiles;

b) determining a third profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject after the treatment, an analyte isolated from a population of phagocytic cells isolated from the subject after the treatment, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after the treatment, an analyte isolated from a population of circulating vesicles isolated from the subject after the treatment, and an analyte isolated from a population of circulating diseased cells isolated from the subject after the treatment;

determining a fourth profile of at least one of the one or more markers from a second control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject after the treatment, an analyte isolated from a population of non-phagocytic cells isolated from the subject after the treatment, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, after the treatment;

identifying a difference between the third and fourth profiles; and

c) identifying a difference between the difference identified in a) and the difference identified in b) wherein the identified difference in c) is indicative of the efficacy of the treatment for said disease or condition in the subject.

14. A method for monitoring the progression or regression of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject at a first time point, an analyte isolated from a population of phagocytic cells isolated from the subject at a first time point, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a first time point, an analyte isolated from a population of circulating vesicles isolated from the subject at a first time point, and an analyte isolated from a population of circulating diseased cells isolated from the subject at a first time point;

determining a second profile of at least one of the one or more markers from a first control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject at a first time point, an analyte isolated from a population of non-phagocytic cells isolated from the subject at a first time point, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, at a first time point;

identifying a difference between the first and second profiles;

b) determining a third profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject at a second time point, an analyte isolated from a population of phagocytic cells isolated from the subject at a second time point, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a second time point, an analyte isolated from a population of circulating vesicles isolated from the subject at a second

time point, and an analyte isolated from a population of circulating diseased cells isolated from the subject at a second time point;

determining a fourth profile of at least one of the one or more markers from a second control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject at a second time point, an analyte isolated from a population of non-phagocytic cells isolated from the subject at a second time point, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, at a second time point;

identifying a difference between the third and fourth profiles; and

c) identifying a difference between the difference identified in a) and the difference identified in b) wherein the identified difference in c) is indicative of the progression or regression of said disease or condition in the subject.

15. A method for identifying a compound capable of ameliorating or treating a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject before administering the compound to the subject, an analyte isolated from a population of phagocytic cells isolated from the subject before administering the compound to the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before administering the compound to the subject, an analyte isolated from a population of circulating vesicles isolated from the subject before administering the compound to the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject before administering the compound to the subject;

determining a second profile of at least one of the one or more markers from a first control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject before administering the compound to the subject, an analyte isolated from a population of non-phagocytic cells isolated from the subject before administering the compound to the subject, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition, before administering the compound to the subject;

identifying a difference between the first and second profiles;

b) determining a third profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject after administering the compound to the subject, an analyte isolated from a population of phagocytic cells isolated from the subject after administering the compound to the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after administering the compound to the subject, an analyte isolated from a population of circulating vesicles isolated from the subject after administering the compound to the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject after administering the compound to the subject;

determining a fourth profile of at least one of the one or more markers from a second control comprising a component selected from the group consisting of: an analyte isolated from a population of phagocytic cells having a DNA content of 2n (=2n phagocytic cells) isolated from the subject after administering the compound to the subject, an analyte isolated from a population of non-phagocytic cells isolated from the subject after administering the compound to the subject, and an analyte isolated from a population of control cells isolated from the subject, wherein the

control cells are substantially free of cells affected by the disease or condition, after administering the compound to the subject;

identifying a difference between the third and fourth profiles; and

c) identifying a difference between the difference identified in a) and the difference identified in b) wherein the identified difference in c) indicates that the compound is capable of ameliorating or treating said disease or condition in the subject.

16. A method for assessing the efficacy of a treatment for a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject before the treatment, a population of phagocytic cells isolated from the subject before the treatment, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before the treatment, an analyte isolated from a population of circulating vesicles isolated from the subject before the treatment, and an analyte isolated from a population of circulating diseased cells isolated from the subject before the treatment;

b) determining a second profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject after the treatment, an analyte isolated from a population of phagocytic cells isolated from the subject after the treatment, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after the treatment, an analyte isolated from a population of circulating vesicles isolated from the subject after the treatment, and an analyte isolated from a population of control cells isolated from the subject, wherein the control cells are substantially free of cells affected by the disease or condition; and

c) identifying a difference between the first profile and the second profile, wherein the identified difference is indicative of the efficacy of the treatment for said disease or condition in the subject.

17. A method for monitoring the progression or regression of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject at a first time point, an analyte isolated from a population of phagocytic cells isolated from the subject at a first time point, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a first time point, an analyte isolated from a population of circulating vesicles isolated from the subject at a first time point, and an analyte isolated from a population of circulating diseased cells isolated from the subject at a first time point;

b) determining a second profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject at a second time point, an analyte isolated from a population of phagocytic cells isolated from the subject at a second time point, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject at a second time point, an analyte isolated from a population of circulating vesicles isolated from the subject at a second time point, and an analyte isolated from a population of circulating diseased cells isolated from the subject at a second time point; and

c) identifying a difference between the first profile and the second profile, wherein the identified difference is indicative of the progression or regression of said disease or condition in the subject.

18. A method for identifying a compound capable of ameliorating or treating a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a first sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject before administering the compound to the subject, an analyte isolated from a population of phagocytic cells isolated from the subject before administering the compound to the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject before administering the compound to the subject, an analyte isolated from a population of circulating vesicles isolated from the subject before administering the compound to the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject before administering the compound to the subject;

b) determining a second profile of one or more markers of the disease or condition from a second sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject after administering the compound to the subject, an analyte isolated from a population of phagocytic cells isolated from the subject after administering the compound to the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject after administering the compound to the subject, an analyte isolated from a population of circulating vesicles isolated from the subject after administering the compound to the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject after administering the compound to the subject; and

c) identifying a difference between the first profile and the second profile, wherein the identified difference indicates that the compound is capable of ameliorating or treating said disease or condition in the subject.

19. A method for diagnosing or aiding in the diagnosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject, a population of phagocytic cells isolated from the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject; and

b) identifying a difference between the first profile and a second profile of at least one of the one or more markers from a repository of said markers of said disease or condition, wherein the difference is indicative of the presence of said disease or condition in the subject.

20. A method for assessing the risk of developing a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject, a population of phagocytic cells isolated from the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject; and

b) identifying a difference between the first profile and a second profile of at least one of the one or more markers from a repository of said markers of said disease or condition , wherein the difference is indicative of the risk of developing said disease or condition in the subject.

21. A method for prognosing or aiding in the prognosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from the subject, a population of phagocytic cells isolated from the subject, a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject; and

b) identifying a difference between the first profile and a second profile of at least one of the one or more markers from a repository of said markers of said disease or condition, wherein the difference is indicative of the prognosis of said disease or condition in the subject.

22. A method for diagnosing or aiding in the diagnosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject, an analyte isolated from a population of phagocytic cells isolated from the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, an analyte isolated from a population of circulating vesicles isolated from the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject; and

b) identifying a difference between the first profile and a second profile of at least one of the one or more markers from a repository of said markers of said disease or condition , wherein the difference is indicative of the presence of said disease or condition in the subject.

23. A method for assessing the risk of developing a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from

the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject, an analyte isolated from a population of phagocytic cells isolated from the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, an analyte isolated from a population of circulating vesicles isolated from the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject; and

b) identifying a difference between the first profile and a second profile of at least one of the one or more markers from a repository of said markers of said disease or condition , wherein the difference is indicative of the risk of developing said disease or condition in the subject.

24. A method for prognosing or aiding in the prognosis of a disease or condition in a subject comprising:

a) determining a first profile of one or more markers of the disease or condition from a sample comprising two or more different components selected from the group consisting of: an analyte isolated from a cell-free bodily fluid isolated from the subject, an analyte isolated from a population of phagocytic cells isolated from the subject, an analyte isolated from a population of phagocytic cells having a DNA content more than 2n (>2n phagocytic cells) isolated from the subject, an analyte isolated from a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject; and

b) identifying a difference between the first profile and a second profile of at least one of the one or more markers from a repository of said markers of said disease or condition , wherein the difference is indicative of the prognosis of said disease or condition in the subject.

25. The method of any one of embodiments 1-3 and 10-12, wherein at least one of the one or more markers is up-regulated or activated in the sample compared to the control.

26. The method of any one of embodiments 1-3 and 10-12, wherein at least one of the one or more markers is down-regulated or inhibited in the sample compared to the control.

27. The method of any one of embodiments 4-6 and 13-15, wherein at least one of the one or more markers is up-regulated or activated in the first sample compared to the first control or in the second sample compared to the second control.

28. The method of any one of embodiments 4-6 and 13-15, wherein at least one of the one or more markers is down-regulated or inhibited in the first sample compared to the first control or in the second sample compared to the second control.

29. The method of any one of embodiments 7-9 and 16-18, wherein at least one of the one or more markers is up-regulated or activated in the first sample compared to the second sample.

30. The method of any one of embodiments 7-9 and 16-18, wherein at least one of the one or more markers is down-regulated or inhibited in the first sample compared to the second sample.

31. The method of any one of embodiments 19-24, wherein at least one of the one or more markers is up-regulated or activated in the sample compared to the repository.

32. The method of any one of embodiments 19-24, wherein at least one of the one or more markers is down-regulated or inhibited in the sample compared to the repository.

33. The method of any one of embodiments 1-32, wherein the first profile or the second profile comprises the absence of at least one of the one or more markers of said disease or condition.

34. The method of any one of embodiments 4-6 and 13-15, wherein the third profile or the fourth profile comprises the absence of at least one of the one or more markers of said disease or condition.

35. The method of any one of embodiments 1-9 and 19-21, wherein, when said method comprises circulating diseased cells, control cells not affected by the disease or condition, phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells, the method further comprises lysing the circulating diseased cells, control cells not affected by the disease or condition, phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells.

36. The method of any one of embodiments 1-9, 19-21, and 35 wherein, when said method comprises circulating diseased cells, control cells not affected by the disease or condition, phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells, the method further comprises extracting at least some of the cellular contents from the circulating diseased cells, control cells not affected by the disease or condition, phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells.

37. The method of any one of embodiments 1-9 and 19-21, 35, and 36, wherein, when said method comprises the cell-free bodily fluid, the method further comprises extracting the one or more markers from the cell-free bodily fluid.

38. The method of any one of embodiments 1-9 and 19-21, wherein, when said method comprises the cell-free bodily fluid, the cell-free bodily fluid comprises a transrenal nucleic acid.

39. The method of any one of embodiments 1-38, wherein at least one of the one or more markers of said disease or condition is present in the circulating diseased cells, cell-free bodily fluid sample, phagocytic cells, or >2n phagocytic cells.

40. The method of any one of embodiments 1-39, wherein at least one of the one or more markers of said disease or condition is not present in the circulating diseased cells, cell-free bodily fluid sample, phagocytic cells, or >2n phagocytic cells.

41. The method of any one of embodiments 1-40, wherein one or more of the circulating diseased cells, control cells not affected by the disease or condition, phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells are enucleated.

42. The method of embodiment 41, wherein the cells are enucleated using physical removal, chemical treatments, photoablation, or ultraviolet irradiation.

43. The method of embodiment 42, wherein the physical removal uses a microneedle, optical tweezer or aspiration.

44. The method of any one of embodiments 19-24, wherein the repository is obtained by data mining.

45. The method of any one of embodiments 1-44, wherein the phagocytic cells, >2n phagocytic cells, or =2n phagocytic cells are neutrophils, macrophages, monocytes, dendritic cells, foam cells, mast cells, eosinophils, keratinocytes, or mixtures thereof.

46. The method of any one of embodiments 1-45, wherein the non-phagocytic cells are T cells, B cells, null cells, basophils, or mixtures thereof.

47. The method of any one of embodiments 1-46, wherein the circulating diseased cells are blood cells, tumor cells, lymphoma cells, fetal cells, apoptotic cells, epithelia cells, endothelial cells, stem cells, progenitor cells, mesenchymal cells, osteoblast cells, osteocytes, hematopoietic stem cells, foam cells, adipose cells, transcervical cells, circulating cardiocytes, circulating fibrocytes, circulating myocytes, circulating cells from kidney, circulating cells from gastrointestinal tract, circulating cells from lung, circulating cells from reproductive organs, circulating cells from central nervous system, circulating hepatic cells, circulating cells from spleen, circulating cells from thymus, circulating cells from thyroid , circulating cells from an endocrine gland, circulating cells from parathyroid, circulating cells from pituitary, circulating cells from adrenal gland, circulating cells from islets of Langerhans, circulating cells from pancreas, circulating cells from hypothalamus, circulating cells from prostate tissues, circulating cells from breast tissues, circulating cells from circulating retinal cells, circulating ophthalmic cells, circulating auditory cells, circulating epidermal cells, circulating cells from the urinary tract, or mixtures thereof.

48. The method of any one of embodiments 1-47, wherein the control cells are normal cells.

49. The method of any one of embodiments 1-48, wherein the control cells are circulating cells.

50. The method of any one of embodiments 1-49, wherein the circulating vesicles are selected from the group consisting of circulating microvesicles, apoptotic bodies, micro-particles, membrane-bound vesicles, multivesicular bodies, nanovesicles, microparticles, and ARRDC-1 mediated microvesicles (ARMM).

51. The method of embodiment 50, wherein the circulating microvesicles are exosomes or urinary exosomes.

52. The method of any one of embodiments 1-51, wherein the cell-free bodily fluid is separated from a bodily fluid.

53. The method of embodiment 52, wherein the bodily fluid is blood, urine, stool, saliva, lymph fluid, cerebrospinal fluid, synovial fluid, cystic fluid, ascites, pleural effusion, fluid obtained from a pregnant woman in the first trimester, fluid obtained from a pregnant woman in the second trimester, fluid obtained from a pregnant woman in the third trimester, maternal blood, amniotic fluid, chorionic villus sample, fluid from a preimplantation embryo, maternal urine, maternal saliva, placental sample, fetal blood, lavage and cervical vaginal fluid, interstitial fluid, buccal swab sample, sputum, bronchial lavage, Pap smear sample, or ocular fluid.

54. The method of any one of embodiments 52 and 53, wherein the cell-free bodily fluid is separated by filtration, centrifugation, flow cytometry, fluorescence activated cell sorting, gradient-based centrifugation, elution, microfluidics, magnetic separation technique, fluorescent-magnetic separation technique, nanostructure, quantum dots, high throughput microscope-based platforms, or a combination thereof.

55. The method of any one of embodiments 52-54, wherein the cell-free bodily fluid is separated by using a substance present in the sample.

56. The method of embodiment 55, wherein the substance is a product of a marker of said disease or condition.

57. The method of any one of embodiments 1-56, wherein the cell-free bodily fluid is plasma or serum.

58. The method of any one of embodiments 1-57, wherein the phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells are isolated from a bodily fluid, tissues, or cells of the subject.

59. The method of embodiment 58, wherein the bodily fluid sample is blood, urine, stool, saliva, lymph fluid, cerebrospinal fluid, synovial fluid, cystic fluid, ascites, pleural effusion, fluid obtained from a pregnant woman in the first trimester, fluid obtained from a pregnant woman in the second trimester, fluid obtained from a pregnant woman in the third trimester, maternal blood, amniotic fluid, chorionic villus sample, fluid from a preimplantation embryo, maternal urine, maternal saliva, placental sample, fetal blood, lavage and cervical vaginal fluid, interstitial fluid, buccal swab sample, sputum, bronchial lavage, Pap smear sample, or ocular fluid.

60. The method of embodiment 58, wherein the cells are white blood cells.

61. The method of any one of embodiments 58-60, wherein the phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells are isolated using antibodies.

62. The method of any one of embodiments 58-60, wherein the phagocytic cells, >2n phagocytic cells, =2n phagocytic cells, or non-phagocytic cells are isolated by flow cytometry, fluorescence activated cell sorting, filtration, gradient-based centrifugation, elution, microfluidics, magnetic separation technique, fluorescent-magnetic separation technique, nanostructure, quantum dots, high throughput microscope-based platforms, or a combination thereof.

63. The method of any one of embodiments 1-62, wherein the one or more markers are nucleic acids, proteins, lipids, carbohydrates, metabolites, or combinations thereof.

64. The method of embodiment 63, wherein the nucleic acids are nucleotides, oligonucleotides, DNAs, RNAs, or DNA-RNA hybrids.

65. The method of embodiment 64, wherein the DNAs are double-stranded DNAs, single-stranded DNAs, multi-stranded DNAs, complementary DNAs, genomic DNAs, or non-coding DNAs.

66. The method of embodiment 64, wherein the RNAs are messenger RNAs (mRNAs), microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small interfering RNAs (siRNAs), heterogeneous nuclear RNAs (hnRNAs), or small hairpin RNAs (shRNAs).

67. The method of embodiment 63, wherein the proteins are amino acids, peptides, enzymes, antigens, antibodies, cytokines, lipoproteins, glycoproteins, or hormones.

68. The method of embodiment 63, wherein the lipids are fatty acids, neutral fats, phosphatides, cholesterol, cholesterol esters, triglycerides, glycolipids, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, polyketides, choline glycerophospholipid, ethanolamine glycerophospholipid, phosphatidylinositol, phosphatidylglycerol, phosphatidylserine, lyso-choline glycerophospholipid, lyso-ethanolamine glycerophospholipid, phosphatidic acid, lyso-phosphatidic acid, sphingomyelin, galactosylceramide, glucosylceramide, free fatty acids, prostaglandins, triacylglycerol, diacylglycerol, monoacylglycerol, acyl-CoA, acylcarnitine, oxysterol, ceramide, cardiolipin, sphingoid base-1-phosphate, shingosine, lyso-sphingomyelin,, gangliosides, plasmalogen, sulfatide, low density lipoproteins (LDLs), very low density lipoproteins (VLDLs), high density lipoproteins (HDLs), sphingoid base-1-phosphates, or derivatives thereof.

69. The method of embodiment 63, wherein the carbohydrates are monosaccharides, disaccharides, polysaccharides, oligosaccharides, or derivatives thereof.

70. The method of embodiment 63, wherein the metabolites are primary metabolites, secondary metabolites, organic metabolites, inorganic metabolites, prostaglandins, hydroxyeicosatetraenoic acids, hydroxyoctadecadienoic acids, steroids, bile acids, vitamins, or derivatives thereof.

71. The method of any one of embodiments 1-3, 10-12, and 19-24, wherein the profile is a nucleic acid profile, a protein profile, a lipid profile, a carbohydrate profile, a metabolite profile, or a combination thereof.

72. The method of embodiment 65, wherein the profile is determined by a qualitative assay, a quantitative assay, or a combination thereof.

73. The method of any one of embodiments 4-9 and 13-18, wherein the first profile or the second profile is a nucleic acid profile, a protein profile, a lipid profile, a carbohydrate profile, a metabolite profile, or a combination thereof.

74. The method of embodiment 65, wherein the first profile or the second profile is determined by a qualitative assay, a quantitative assay, or a combination thereof.

75. The method of any one of embodiments 4-6 and 74, wherein the third profile or the fourth profile is a nucleic acid profile, a protein profile, a lipid profile, a carbohydrate profile, a metabolite profile, or a combination thereof.

76. The method of embodiment 75, wherein the third profile or the fourth profile is determined by a qualitative assay, a quantitative assay, or a combination thereof.

77. The method of any one of embodiments 72, 66, and 76, wherein the quantitative assay uses sequencing, targeted sequencing, single molecule real-time sequencing, electron microscopy-based sequencing, transistor-mediated sequencing, direct sequencing, random shotgun sequencing, Sanger dideoxy termination sequencing, exon sequencing, whole-genome sequencing, sequencing by hybridization, pyrosequencing, capillary electrophoresis, gel electrophoresis, duplex sequencing, cycle sequencing, single-base extension sequencing, solid-phase sequencing, high-throughput sequencing, massively parallel signature sequencing, emulsion PCR, multiplex PCR, co-amplification at lower denaturation temperature-PCR (COLD-PCR), sequencing by reversible dye terminator, paired-end sequencing, near-term sequencing, exonuclease sequencing, sequencing by ligation, short-read sequencing, single-molecule sequencing, sequencing-by-synthesis, real-time sequencing, reverse-terminator sequencing, nanopore sequencing, 454 sequencing, Solexa Genome Analyzer sequencing, SOLiD® sequencing, MS-PET sequencing,

mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), polymerase chain reaction (PCR) analysis, co-amplification at lower denaturation temperature-PCR (COLD-PCR), multiplex PCR, quantitative PCR, real-time PCR, fluorescence assay, colorimetric assay, chemiluminescent assay, or a combination thereof.

78. The method of embodiment 77, wherein the nucleic acid profile is a genotypic profile, a single nucleotide polymorphism profile, a gene mutation profile, a gene copy number profile, a DNA methylation profile, a DNA acetylation profile, a chromosome dosage profile, a gene expression profile, or a combination thereof.

79. The method of embodiment 77, wherein the nucleic acid profile is determined by polymerase chain reaction (PCR) analysis, sequencing analysis, electrophoretic analysis, restriction fragment length polymorphism (RFLP) analysis, Northern blot analysis, quantitative PCR, reverse-transcriptase-PCR analysis (RT-PCR), allele-specific oligonucleotide hybridization analysis, comparative genomic hybridization, heteroduplex mobility assay (HMA), single strand conformational polymorphism (SSCP), denaturing gradient gel electrophisis (DGGE), RNAase mismatch analysis, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), surface plasmon resonance, Southern blot analysis, in situ hybridization, fluorescence in situ

hybridization (FISH), chromogenic in situ hybridization (CISH), immunohistochemistry (IHC), microarray, comparative genomic hybridization, karyotyping, multiplex ligation-dependent probe amplification (MLPA), Quantitative Multiplex PCR of Short Fluorescent Fragments (QMPSF), microscopy, methylation specific PCR (MSP) assay, HpaII tiny fragment Enrichment by Ligation-mediated PCR (HELP) assay, radioactive acetate labeling assays, colorimetric DNA acetylation assay, chromatin immunoprecipitation combined with microarray (ChIP-on-chip) assay, restriction landmark genomic scanning, Methylated DNA immunoprecipitation (MeDIP), molecular break light assay for DNA adenine methyltransferase activity, chromatographic separation, methylation-sensitive restriction enzyme analysis, bisulfite-driven conversion of non-methylated cytosine to uracil, co-amplification at lower denaturation temperature-PCR (COLD-PCR), multiplex PCR, methyl-binding PCR analysis, or a combination thereof.

80. The method of embodiment 77, wherein the nucleic acid profile is determined by a sequencing technique selected from the group consisting of targeted sequencing, single molecule real-time sequencing, exon sequencing, electron microscopy-based sequencing, transistor-mediated sequencing, direct sequencing, random shotgun sequencing, Sanger dideoxy termination sequencing, whole-genome sequencing, sequencing by hybridization, pyrosequencing, capillary electrophoresis, gel electrophoresis, duplex sequencing, cycle sequencing, single-base extension sequencing, solid-phase sequencing, high-throughput sequencing, massively parallel signature sequencing, emulsion PCR, co-amplification at lower denaturation temperature-PCR (COLD-PCR), multiplex PCR, sequencing by reversible dye terminator, paired-end sequencing, near-term sequencing, exonuclease sequencing, sequencing by ligation, short-read sequencing, single-molecule sequencing, sequencing-by-synthesis, real-time sequencing, reverse-terminator sequencing, nanopore sequencing, 454 sequencing, Solexa Genome Analyzer sequencing, SOLiD® sequencing, MS-PET sequencing, mass spectrometry, and a combination thereof.

81. The method of embodiment 77, wherein the protein profile is a protein expression profile, a protein activation profile, or a combination thereof.

82. The method of embodiment 77, wherein the protein profile is determined by an immunohistochemistry assay, an enzyme-linked immunosorbent assay (ELISA), in situ hybridization, chromatography, liquid chromatography, size exclusion chromatography, high performance liquid chromatography (HPLC), gas chromatography, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), radioimmunoassays, microscopy, microfluidic chip-based assays, surface plasmon resonance, sequencing, Western blotting assay, or a combination thereof.

83. The method of embodiment 77, wherein the protein activation profile comprises determining a phosphorylation state, an ubiquitination state, a myristoylation state, a conformational state, or a combination thereof of the one or more markers.

84. The method of embodiment 77, wherein the lipid profile is determined by chromatography, liquid chromatography, size exclusion chromatography, high performance liquid chromatography (HPLC), gas chromatography, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), radioimmunoassays, microfluidic chip-based assay, detection of fluorescence, detection of chemiluminescence, or a combination thereof.

85. The method of embodiment 77, wherein the carbohydrate profile is determined by chromatography, liquid chromatography, size exclusion chromatography, high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), liquid chromatography, gas chromatography, fluorescent assay, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), radioimmunoassay, microfluidic chip-based assay, detection of fluorescence, detection of chemiluminescence, or a combination thereof.

86. The method of any one of embodiments 1-85, wherein the subject has at least two diseases or conditions.

87. The method of embodiment 86, wherein the subject has at least one prenatal or pregnancy-related disease or condition

88. The method of any one of embodiments 1-87, wherein the subject is a mammal.

89. The method of embodiment 89, wherein the mammal is a human.

90. The method of any one of embodiments 1-89, wherein the disease or condition is a cardiovascular disease or condition, a kidney-associated disease or condition, a prenatal or pregnancy-related disease or condition, a neurological or neuropsychiatric disease or condition, an autoimmune or immune-related disease or condition, a cancer, an infectious disease or condition, a mitochondrial disorder, a respiratory-gastrointestinal tract disease or condition, a reproductive disease or condition, an ophthalmic disease or condition, a musculo-skeletal disease or condition, or a dermal disease or condition.

91. The method of any one of embodiments 1-90, wherein the difference is greater than a 1-fold difference.

92. The method of embodiment 91, wherein the difference is at least 1.05-fold, 1.1-fold, 1.2-fold, 1.3-fold, 1.4-fold, 1.5-fold, 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold difference.

93. The method of any one of embodiments 4-6 and 13-15, wherein the second profile and the fourth profile are the same.

94. A method for identifying one or more markers for a disease or condition comprising:

a) determining a first profile of analytes from a sample comprising a cell-free bodily fluid from a subject having said disease or condition and a population of phagocytic cells, or a population of >2n phagocytic cells, from a subject having said disease or condition;

determining a second profile of analytes from a population of =2n phagocytic cells, or a population of non-phagocytic cells, from the subject having said disease or condition;

identifying a set of differences between the first and second profiles, wherein the first set of differences is specific to the first profile relative to the second profile;

b) determining a third profile of analytes from a sample comprising a cell-free bodily fluid from a subject having said disease or condition and a population of phagocytic cells, or a population of >2n phagocytic cells, from a control subject not having said disease or condition;

determining a fourth profile of analytes from a population of =2n phagocytic cells, or a population of non-phagocytic cells, from the control subject not having said disease or condition;

identifying a set of differences between the third and fourth profiles, wherein the second set of differences is specific to the third profile relative to the fourth profile; and

c) identifying one or more analytes specific to the set of differences identified in a) relative to the set of differences identified in b), the identified analytes in c) being markers of said disease or condition.

95. A method for identifying one or more markers for a disease or condition comprising:

a) determining a first profile of analytes from a sample comprising a cell-free bodily fluid from a subject having said disease or condition and a population of phagocytic cells, or a population of >2n phagocytic cells, from a subject having said disease or condition;

b) comparing the first profile to a second profile derived from a repository of analytes from a control subject not having said disease or condition;

c) identifying a set of differences between the first and second profiles, wherein the set of differences is specific to the first profile relative to the second profile; and

d) identifying one or more analytes specific to the set of differences, the identified analytes being markers of said disease or condition.

96. The method of embodiment 95, further comprising:

a) obtaining a fifth profile of analytes from cells or tissues affected by said disease or condition in the subject having said disease or condition;

obtaining a sixth profile of analytes from cells or tissues not affected by said disease or condition in the subject having said disease or condition;

identifying a set of differences between the fifth and sixth profiles, wherein the set of differences is specific to the fifth profile relative to the sixth profile; and

b) identifying at least one of the one or more markers of c) present in the set of differences identified in d).

97. The method of any one of embodiments 90-96, further comprising extracting the one or more markers from the cell-free bodily fluid.

98. The method of any one of embodiments 90-97, further comprising lysing the phagocytic cells, the >2n phagocytic cells, the =2n phagocytic cells, or the non-phagocytic cells before a).

99. The method of any one of embodiments 90-98, further comprising extracting at least some of the cellular contents from the phagocytic cells, the >2n phagocytic cells, the =2n phagocytic cells, or the non-phagocytic cells before a).

100. The method of any one of embodiments 90-99, wherein the phagocytic cells or the >2n phagocytic cells comprise viable diseased cells, dead diseased cells, apoptotic diseased cells, circulating tumor cells, infectious agents, fetal cells, trophoblasts, or fragments thereof.

101. The method of any one of embodiments 90-100, wherein at least one of the one or more markers of said disease or condition is present in the cell-free bodily fluid sample, the phagocytic cells, or the >2n phagocytic cells.

102. The method of any one of embodiments 90-101, wherein at least one of the one or more markers of said disease or condition is not present in the cell-free bodily fluid sample, the phagocytic cells, or the >2n phagocytic cells.

103. The method of any one of embodiments 90-102, further comprising comparing the identified difference of c) to a repository of one or more known markers of said disease or condition.

104. The method of embodiment 97, wherein the repository is obtained by data mining.

105. The method of any one of embodiments 90-104, wherein the phagocytic cells, >2n phagocytic cells, or =2n phagocytic cells are neutrophils, macrophages, monocytes, dendritic cells, foam cells, mast cells, eosinophils, keratinocytes, or mixtures thereof.

106. The method of any one of the embodiments 90-105, wherein the non-phagocytic cells are T cells, B cells, null cells, basophils, or mixtures thereof

107. The method of any one of embodiments 90-106, wherein the cell-free bodily fluid is separated from a bodily fluid.

108. The method of embodiment 107, wherein the bodily fluid sample is blood, urine, stool, saliva, lymph fluid, cerebrospinal fluid, synovial fluid, cystic fluid, ascites, pleural effusion, fluid obtained from a pregnant woman in the first trimester, fluid obtained from a pregnant woman in the second trimester, fluid obtained from a pregnant woman in the third trimester, maternal blood, amniotic fluid, chorionic villus sample, fluid from a preimplantation embryo, maternal urine, maternal saliva, placental sample, fetal blood, lavage and cervical vaginal fluid, interstitial fluid, or ocular fluid.

109. The method of any one of embodiments 107 and 108, wherein the cell-free bodily fluid sample is separated by filtration, centrifugation, flow cytometry, fluorescence activated cell sorting, gradient-based centrifugation, elution, microfluidics, magnetic separation technique, fluorescent-magnetic separation technique, nanostructure, quantum dots, high throughput microscope-based platforms, or a combination thereof.

110. The method of embodiment 109, wherein the cell-free bodily fluid sample is separated by using a substance present in the sample.

111. The method of any one of embodiments 90-110, wherein the one or more markers are nucleic acids, proteins, lipids, carbohydrates, metabolites, or combinations thereof.

112. The method of embodiment 111, wherein the nucleic acids are nucleotides, oligonucleotides, DNAs, RNAs, or DNA-RNA hybrids.

113. The method of embodiment 112, wherein the DNAs are double-stranded DNAs, single-stranded DNAs, multi-stranded DNAs, complementary DNAs, genomic DNAs, or non-coding DNAs.

114. The method of embodiment 112, wherein the RNAs are messenger RNAs (mRNAs), microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small interfering RNAs (siRNAs), heterogeneous nuclear RNAs (hnRNAs), or small hairpin RNAs (shRNAs).

115. The method of embodiment 111, wherein the proteins are amino acids, peptides, enzymes, antigens, antibodies, cytokines, lipoproteins, glycoproteins, or hormones.

116. The method of embodiment 111, wherein the lipids are fatty acids, neutral fats, phosphatides, cholesterol, cholesterol esters, triglycerides, glycolipids, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, polyketides, choline glycerophospholipid, ethanolamine glycerophospholipid, phosphatidylinositol, phosphatidylglycerol, phosphatidylserine, lyso-choline glycerophospholipid, lyso-ethanolamine glycerophospholipid, phosphatidic acid, lyso-phosphatidic acid, sphingomyelin, galactosylceramide, glucosylceramide, free fatty acids, prostaglandins, triacylglycerol, diacylglycerol, monoacylglycerol, acyl-CoA, acylcarnitine, oxysterol, ceramide, cardiolipin, sphingoid base-1-phosphate, shingosine, lyso-sphingomyelin,, gangliosides, plasmalogen, sulfatide, low density lipoproteins (LDLs), very low density lipoproteins (VLDLs), high density lipoproteins (HDLs), sphingoid base-1-phosphates, or derivatives thereof.

117. The method of embodiment 111, wherein the carbohydrates are monosaccharides, disaccharides, polysaccharides, oligosaccharides, or derivatives thereof.

118. The method of embodiment 111, wherein the metabolites are primary metabolites, secondary metabolites, organic metabolites, inorganic metabolites, prostaglandins, hydroxyeicosatetraenoic acids, hydroxyoctadecadienoic acids, steroids, bile acids, vitamins, or derivatives thereof.

119. The method of any one of embodiments 90-118, wherein the profile is a nucleic acid profile, a protein profile, a lipid profile, a carbohydrate profile, a metabolite profile, or a combination thereof.

120. The method of embodiment 119, wherein the profile is determined by a qualitative assay, a quantitative assay, or a combination thereof.

121. The method of embodiment 120, wherein the quantitative assay uses sequencing, targeted sequencing, single molecule real-time sequencing, exon sequencing, electron microscopy-based sequencing, transistor-mediated sequencing, direct sequencing, random shotgun sequencing, Sanger dideoxy termination sequencing, whole-genome sequencing, sequencing by hybridization, pyrosequencing, capillary electrophoresis, gel electrophoresis, duplex sequencing, cycle sequencing, single-base extension sequencing, solid-phase sequencing, high-throughput sequencing, massively parallel signature sequencing, emulsion PCR, co-amplification at lower denaturation temperature-PCR (COLD-PCR), multiplex PCR, sequencing by reversible dye terminator, paired-end sequencing, near-term sequencing, exonuclease sequencing, sequencing by ligation, short-read sequencing, single-molecule sequencing, sequencing-by-synthesis, real-time sequencing, reverse-terminator sequencing, nanopore sequencing, 454 sequencing, Solexa Genome Analyzer sequencing, SOLiD® sequencing, MS-PET sequencing, mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass

spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), polymerase chain reaction (PCR) analysis, quantitative PCR, real-time PCR, fluorescence assay, colorimetric assay, chemiluminescent assay, or a combination thereof.

122. The method of embodiment 119, wherein the nucleic acid profile is a genotypic profile, a single nucleotide polymorphism profile, a gene mutation profile, a gene copy number profile, a DNA methylation profile, a DNA acetylation profile, a chromosome dosage profile, a gene expression profile, or a combination thereof.

123. The method of embodiment 119, wherein the nucleic acid profile is determined by polymerase chain reaction (PCR) analysis, sequencing analysis, electrophoretic analysis, restriction fragment length polymorphism (RFLP) analysis, Northern blot analysis, quantitative PCR, reverse-transcriptase-PCR analysis (RT-PCR), allele-specific oligonucleotide hybridization analysis, comparative genomic hybridization, heteroduplex mobility assay (HMA), single strand conformational polymorphism (SSCP), denaturing gradient gel electrophisis (DGGE), RNAase mismatch analysis, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), surface plasmon resonance, Southern blot analysis, in situ hybridization, fluorescence in situ hybridization (FISH), chromogenic in situ hybridization (CISH), immunohistochemistry (IHC), microarray, comparative genomic hybridization, karyotyping, multiplex ligation-dependent probe amplification (MLPA), Quantitative Multiplex PCR of Short Fluorescent Fragments (QMPSF), microscopy, methylation specific PCR (MSP) assay, HpaII tiny fragment Enrichment by Ligation-mediated

PCR (HELP) assay, radioactive acetate labeling assays, colorimetric DNA acetylation assay, chromatin immunoprecipitation combined with microarray (ChIP-on-chip) assay, restriction landmark genomic scanning, Methylated DNA immunoprecipitation (MeDIP), molecular break light assay for DNA adenine methyltransferase activity, chromatographic separation, methylation-sensitive restriction enzyme analysis, bisulfite-driven conversion of non-methylated cytosine to uracil, co-amplification at lower denaturation temperature-PCR (COLD-PCR), multiplex PCR, methyl-binding PCR analysis, or a combination thereof.

124. The method of embodiment 119, wherein the nucleic acid profile is determined by a sequencing technique selected from the group consisting of targeted sequencing, single molecule real-time sequencing, exon sequencing, electron microscopy-based sequencing, transistor-mediated sequencing, direct sequencing, random shotgun sequencing, Sanger dideoxy termination sequencing, whole-genome sequencing, sequencing by hybridization, pyrosequencing, capillary electrophoresis, gel electrophoresis, duplex sequencing, cycle sequencing, single-base extension sequencing, solid-phase sequencing, high-throughput sequencing, massively parallel signature sequencing, emulsion PCR, co-amplification at lower denaturation temperature-PCR (COLD-PCR), multiplex PCR, sequencing by reversible dye terminator, paired-end sequencing, near-term sequencing, exonuclease sequencing, sequencing by ligation, short-read sequencing, single-molecule sequencing, sequencing-by-synthesis, real-time sequencing, reverse-terminator sequencing, nanopore sequencing, 454 sequencing, Solexa Genome Analyzer sequencing, SOLiD® sequencing, MS-PET sequencing, mass spectrometry, and a combination thereof.

125. The method of embodiment 119, wherein the protein profile is a protein expression profile, a protein activation profile, or a combination thereof.

126. The method of embodiment 119, wherein the protein profile is determined by an immunohistochemistry assay, an enzyme-linked immunosorbent assay (ELISA), in situ hybridization, chromatography, liquid chromatography, size exclusion chromatography, high performance liquid chromatography (HPLC), gas

chromatography, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), radioimmunoassays, microscopy, microfluidic chip-based assays, surface plasmon resonance, sequencing, Western blotting assay, or a combination thereof.

127. The method of embodiment 119, wherein the protein activation profile comprises determining a phosphorylation state, an ubiquitination state, a myristoylation state, a conformational state, or a combination thereof of the one or more markers.

128. The method of embodiment 119, wherein the lipid profile is determined by chromatography, liquid chromatography, size exclusion chromatography, high performance liquid chromatography (HPLC), gas chromatography, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), radioimmunoassays, microfluidic chip-based assay, detection of fluorescence, detection of chemiluminescence, or a combination thereof.

129. The method of embodiment 119, wherein the carbohydrate profile is determined by chromatography, liquid chromatography, size exclusion chromatography, high performance anion exchange chromatography with pulsed

amperometric detection (HPAEC-PAD), liquid chromatography, gas chromatography, fluorescent assay, mass spectrometry, tandem mass spectrometry, matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, electrospray ionization (ESI) mass spectrometry, surface-enhanced laser deorption/ionization-time of flight (SELDI-TOF) mass spectrometry, quadrupole-time of flight (Q-TOF) mass spectrometry, atmospheric pressure photoionization mass spectrometry (APPI-MS), Fourier transform mass spectrometry (FTMS), matrix-assisted laser desorption/ionization-Fourier transform-ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry, secondary ion mass spectrometry (SIMS), radioimmunoassay, microfluidic chip-based assay, detection of fluorescence, detection of chemiluminescence, or a combination thereof.

130. The method of any one of embodiments 90-129, wherein the subject is a mammal

131. The method of embodiment 130, wherein the mammal is a human.

132. The method of any one of embodiments 90-131, wherein the disease or condition is a cardiovascular disease or condition, a kidney-associated disease or condition, a prenatal or pregnancy-related disease or condition, a neurological or neuropsychiatric disease or condition, an autoimmune or immune-related disease or condition, a cancer, an infectious disease or condition, a mitochondrial disorder, a respiratory-gastrointestinal tract disease or condition, a reproductive disease or condition, an ophthalmic disease or condition, a musculo-skeletal disease or condition, or a dermal disease or condition.

133. The method of any one of embodiments 90-132, wherein the difference is greater than a 1-fold difference.

134. The method of embodiment 133, wherein the difference is at least 1.05-fold, 1.1-fold, 1.2-fold, 1.3-fold, 1.4-fold, 1.5-fold, 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold difference.

135. The method of any one of embodiments 1-93, further comprising determining at least one diagnostic parameter of said disease or condition.

136. The method of embodiment 135, wherein the diagnostic parameter is determined by physical inspection, visual inspection, biopsy, scanning, histology, radiology, imaging, ultrasound, use of a commercial kit, genetic testing, immunological testing, analysis of bodily fluids, or monitoring neural activity.

137. The method of any one of the embodiments 1-93 and 135-136, wherein the one or more markers comprise at least one gene selected from the group consisting of AKT2, BAK1, EGFR, ERBB2, ETS2, FOS, JUN, MAP2K1, MMP2, PDGFB, RB1, SERPINB2, SNCG, and SPP1.

138. The method of any one of the embodiments 1-93 and 135-137, wherein the one or more markers comprise at least one gene selected from the group consisting of AKT1, AKT2, BAK2, CDC25A, E2F1, EGFR, ERBB2, FOS, JUN, MAP2K1, MMP2, NFKB1, PDGFB, PIK3R1, PNN, RB1, SERPINB2, SERPINB5, SNCG, SPP1, TERT, TIMP3, and TP53.

139. The method of any one of the embodiments 1-93 and 135-138, wherein the one or more markers comprise at least one gene selected from the group consisting of CASP8, CASP9, COL18A1, ETS2, HTATIP2, MMP9, SRC, and TWIST1.

140. The method of any one of the embodiments 1-93 and 135-139, wherein the one or more markers comprise at least one gene selected from the group consisting of AKT1, APAF1, ATM, CDC25A, CDKN1A, ETS2, FOS, IL8, ITGA4, ITGA6, ITGAV, JUN, MAP2K1, NFKBIA, PLAU, PLAUR, RAF1, SERPINB2, SYK, TIMP1, TNF, TNFRSF10B, and TNFRSF1A.

141. The method of any one of the embodiments 1-93 and 135-140, wherein the one or more markers comprise at least one gene selected from the group consisting of ACP2, AK2, AKT3, ARL5B, ATP2B3, BGN, BRAF, BTG2, CAMKK2, CAPG, CAPN12, CPLX2, DENND5A, DNA2, FAM104A, FNIP1, GFRA4, GLUD1, GNAQ, GP1BB, HNRPLL, HOXA2, HPS3, INPP4A, ITGAV, KLHL23, LANCL2, LYPD6, MAPKAPK3, MEF2A (includes, EG:4205), MEF2C, NVL, PCYT1A, PGLYRP4, PLOD1, PPP1CB, PRKAB2, PROS1, PTPRE, RASA4

(includes,EG:10156), RBMS2, RBPJ, STAT5B, THBS1, TRIB1, TRIM2, TSPAN6, and ZDHHC21.

142. The method of any one of the embodiments 1-93 and 135-141, wherein the one or more markers comprise at least one gene selected from the group consisting of B4GALT5, BOP1, CCL2, CCL3, CCL3L1, CCRL2, CD83, CLEC4G, CLIC4, CTSC, CTSO, CXCL10, FCGR3A, FPR3, HBA1, HBB, LRMP, MAP1LC3B2, MS4A4A, MSR1, MYADML, NID1, PF4, PION, RNF217, SAMD9L, SERPING1, and SPARC.

143. The method of any one of the embodiments 1-93 and 135-142, wherein the one or more markers comprise at least one gene selected from the group consisting of ACOT9, AMPD2, ARHGAP15, BATF2, C3AR1, C5orf41, CCL3, CCL3L1, CD63, CHST11, CHSY1, CLEC4G, CTSZ, CXorf21, CYTH4, CYTIP, DLEU2, DNAJA1, DOCK8, DTX3L, DUSP6, EPSTI1, ERF, F2RL1, FYB, GABRB2, GBP5, GLRX, GNB4, ICAM1, IFI35, IFIH1, IFNAR2, IL1R1, IRF1, ITGA5, LAP3, LAPTM5, LCP2, MAP1LC3B, MAP1LC3B2, MICAL2, MT1DP, MT1JP, MT1M, MT2A, MYADML, NEK6, NINJ2, NNMT, NT5C3L, NUB1, PDE4B, PLOD1, PML, PRKCB, PSMB9, RCN3, RGS4, RNASE6, RTP4, SAMD9L, SEL1L, SERPING1, SETX, SIGLEC10, SKIL, SLC7A7, SNORA21, SP100, SP110, SP140, SSFA2, STAT2, STK17B, STK3, TDRD7, TMCC1, TMPRSS11E2, TNFRSF1B, TPM1, TRIM21,TXNDC4, UBE2L6, UBE2W, USP18, VAV1, WARS, WIPF1, and WIPI1.

144. The method of any one of the embodiments 1-93 and 135-143, wherein the one or more markers comprise at least one gene selected from the group consisting of ADAR, ADM, ALAS1, ANKRD22, ARHGAP27, B3GNT5, BCL10, C12orf35, C15orf29, C2orf59, CD177, CEACAM1, CPEB2, DDX58, F2RL1, GDPD3, GNAI3, HIST2H3A, HIST2H3D, HIST2H4A, HMGCR, HSPA6, HSPC159, IL4R, IMPA2, KPNB1, KREMEN1, KRT23, LDLR, LOC100130904, LTB4R, MAEA, MARK2, MBOAT2, MPZL3, N4BP1, NBEAL2, NMI, NPEPPS, PARP14, PGM2, PPIF, PXN, RALBP1, ROD1, RPS6KA1, S100P, SERTAD2, SLC9A1, SLPI, SP110, SPINT1, ST14, TBC1D3, TNFRSF9, TRIM21, UPP1, VPS24, ZBTB34, and ZNF256.

145. The method of any one of the embodiments 1-93 and 135-144, wherein the one or more markers comprise at least one or more of the markers identified by the methods of any one the embodiments 94-134.

146. A kit comprising a plurality of marker detection agents that detect at least one or more of the markers identified by the methods of any one of the embodiments 94-134.

147. A method of treating or preventing a disease or condition in a subject comprising administering to said subject a composition comprising a compound identified by the method of any one of embodiments 6 and 15.

148. The method of any one of the embodiments 1-93 and 135-145, wherein the circulating diseased cells are infected by an infectious agent.

149. The method of embodiment 148, wherein the infectious agent is a virus, bacteria, fungus, parasite, protozoan, infectious protein or microorganism.

150. The method of any one of embodiments 1-93, 135-145, and 148-149, wherein, when said method comprises phagocytic cells or >2n phagocytic cells, said phagocytic cells or >2n phagocytic cells comprise a transrenal nucleic acid.

Brief Description of the Drawings

[0008] Figure 1 depicts a set of potentially informative markers and the relative percentage of reads of associated allelic frequencies in maternal reference samples of CD2-positive peripheral blood cells and combination samples. Along the x-axis are the markers and along the y-axis are the percent reads for the different alleles. Triangle: percent of maternal sample reference allele; circle: percent of maternal sample alternate allele; square: percent of combination sample reference allele; and diamond: percent of combination sample alternate allele.

[0009] Figure 2 depicts a set of potentially informative markers and the relative percentage of reads of associated allelic frequencies in maternal reference samples and combination samples. Along the x-axis are the markers and along the y-axis are the percent reads for the different alleles. Triangle: percent of maternal sample

reference allele; circle: percent of maternal sample alternate allele; square: percent of combination sample reference allele; and diamond: percent of combination sample alternate allele.

[0010] Figure 3 depicts a set of potentially informative markers and the relative percentage of reads of associated allelic frequencies as compared between a circulating cell-free fluid (ccff) sample and a maternal reference sample. Along the x-axis are the markers and along the y-axis are the percent reads for the different alleles. Triangle: reference allele; circle: alternative allele.

[0011] Figure 4 depicts a set of potentially informative markers and the relative percentage of reads of associated allelic frequencies in a T-cell sample. Along the x-axis are the markers and along the y-axis are the percent reads for the different alleles. Triangle: reference allele; circle: alternative allele.

[0012] Figure 5 depicts a set of potentially informative markers and the relative percentage of reads of associated allelic frequencies in a combination sample of a circulating cell-free fluid (ccff) plus 1% monocyte sample. Along the x-axis are the markers and along the y-axis are the percent reads for the different alleles. Triangle: reference allele; circle: alternative allele.

[0013] Figure 6 depicts the emergence of informative markers in a circulating cell-free fluid fraction of the maternal blood compared to the lack of informative markers in the maternal T-cells. Along the x-axis are the markers and along the y-axis are the percent reads for the different alleles.

[0014] Figure 7 depicts the emergence of the informative markers in a combination sample containing cell-free bodily fluid plus monocytes from maternal blood compared to a lack of informative markers in the maternal reference sample containing T-cells. Along the x-axis are the markers and along the y-axis are the percent reads for the different alleles.

[0015] Figure 8 depicts an exemplary informative marker in a maternal only sample and a combination sample. Triangle: reference allele; circle: alternative allele.

[0016] Figure 9 depicts an exemplary informative marker in a maternal only sample and a combination sample. Triangle: reference allele; circle: alternative allele.

[0017] Figure 10 depicts an exemplary non-informative marker in a maternal only sample and a combination sample. Triangle: reference allele; circle: alternative allele.

Detailed Description of the Invention

[0018] Unless otherwise defined herein, scientific and technical terms used in this application shall have the meanings that are commonly understood by those of ordinary skill in the art. Generally, nomenclature used in connection with, and techniques of, cell and tissue culture, molecular biology, cell and cancer biology, neurobiology, neurochemistry, virology, immunology, microbiology, pharmacology, genetics and protein and nucleic acid chemistry, described herein, are those well known and commonly used in the art.

[0019] All of the above, and any other publications, patents and published patent applications referred to in this application are specifically incorporated by reference herein. In case of conflict, the present specification, including its specific definitions, will control.

[0020] Throughout this specification, the word "comprise" or variations such as "comprises" or "comprising" will be understood to imply the inclusion of a stated integer (or components) or group of integers (or components), but not the exclusion of any other integer (or components) or group of integers (or components).

[0021] The singular forms "a," "an," and "the" include the plurals unless the context clearly dictates otherwise.

[0022] The term "including" is used to mean "including but not limited to". "Including" and "including but not limited to" are used interchangeably.

[0023] A "patient", "subject", or "individual" are used interchangeably and refer to either a human or a non-human animal. These terms include mammals, such as humans, primates, livestock animals (e.g., bovines, porcines), companion animals (e.g., canines, felines) and rodents (e.g., mice and rats).

[0024] As used herein, a control subject refers to any individual that has not been diagnosed as having the disease or condition being assayed. The terms "normal control", "healthy control", and "not-diseased cells" likewise mean a sample (e.g., cells, serum, tissue) taken from a source (e.g., subject, control subject, cell line) that does not have the condition or disease being assayed and therefore may be used to determine the baseline for the condition or disorder being measured. It is also understood that the control subject, normal control, and healthy control, include data obtained and used as a standard, i.e. it can be used over and over again for multiple different subjects. In other words, for example, when comparing a subject sample to a control sample, the data from the control sample could have been obtained in a different set of experiments, for example, it could be an average obtained from a number of healthy subjects and not actually obtained at the time the data for the subject was obtained.

[0025] The term "diagnosis" as used herein refers to methods by which the skilled artisan can estimate and/or determine whether or not a patient is suffering from a given disease or condition. The skilled artisan often makes a diagnosis on the basis of one or more diagnostic indicators, e.g., a marker, the presence, absence, amount, or change in amount of which is indicative of the presence, severity, or absence of the condition.

[0026] The term "prognosis" as used herein refers to is used herein to refer to the likelihood of a disease or condition progression, including recurrence of a disease or condition.

[0027] The disclosure of the International Applications PCT/US09/31395, PCT/US11/45009, PCT/US11/44969, PCT/US11/44973, PCT/US11/44991, PCT/US11/45002, PCT/US11/44996, and PCT/US11/45018 are incorporated herein by reference for all purposes.

[0028] Each embodiment described herein may be combined with any other embodiment described herein.

Description of Methods of the Invention

[0029] By using a profile derived from a combination of sources, the skilled worker is able to capture data that is normally lost in the process of separating components out of a sample for testing. At the same time, the analytical components being combined are enriched for the markers used. Thus, the methods of the invention do not introduce unnecessary "noise" into the signal. Embodiments using a subject-specific profile comparison also eliminate the dependence on a population-derived average profile for a particular disease or condition. Using a population-derived average profile may introduce error into the detection or diagnosis of a particular disease or condition in the subject. Therefore, methods of this invention allow detection, diagnosis, and treatment to be personalized to the individual.

[0030] The methods of this invention have high specificity, sensitivity, and accuracy, and are capable of detecting disease or condition-specific markers present within a bodily fluid sample, cells or tissues. In some embodiments, the methods of this invention have improved specificity, sensitivity, and accuracy compared to current methods. The methods of the invention also reduce the problem of losing valuable signal when, for example, a patient sample is collected and separated into individual components, of which only one is chosen for testing. Accordingly, in certain aspects, the invention provides non-invasive assays for the early detection of a disease or condition, i.e., before the disease can be diagnosed by conventional diagnostic techniques, e.g., imaging techniques, and, therefore, provide a foundation for improved decision-making relative to the needs and strategies for intervention, prevention, and treatment of individuals with such disease or condition.

[0031] The present invention provides methods for diagnosing or aiding in the diagnosis of a disease or condition by comparing profiles (e.g., gene/protein/lipid/carbohydrate expression profiles, genotypes, gene copy number, gene dosage, DNA methylation, etc.) of disease or condition-associated markers (e.g., nucleic acids, proteins, lipids, carbohydrates, metabolites). One profile used for the comparison may be a profile from a sample comprising combinations of two or more different components (e.g., cell-free bodily fluids, phagocytic cells, circulating vesicles, and circulating diseased cells) from a subject. Alternatively, profile used for the comparison may be a profile from a sample comprising analytes isolated from,

e.g., cell-free bodily fluids, phagocytic cells, phagocytic cells having a DNA content more than 2n (>2n phagocytic cells), circulating vesicles, and circulating diseased cells from a subject. For example, a sample comprising analytes from phagocytes and a cell-free bodily fluid could be generated by isolating analytes (e.g., nucleic acids or proteins) from phagocytes (e.g., by lysing the cells and using affinity-based techniques to isolate the analytes) and analytes (e.g., nucleic acids or proteins) from a cell-free bodily fluid, and combining the analytes to create a sample useful in creating a profile. For ease of reference, a sample comprising combinations of two or more different components (e.g., cell-free bodily fluids, phagocytic cells, >2n phagocytic cells, circulating vesicles, and circulating diseased cells, or analytes isolated therefrom), may be referred to herein as a "combination sample." In some embodiments, the combination sample may comprise a lysate, or a fraction or portion of a lysate, of a population of phagocytic cells, a population of >2n phagocytic cells, a population of circulating vesicles, or a population of circulating diseased cells in place of the population of phagocytic cells, the population of >2n phagocytic cells, the population of circulating vesicles, or the population of circulating diseased cells, respectively. In some embodiments, the combination sample may comprise an analyte isolated from a lysate, or a fraction or portion of an analyte isolated from a lysate, of a population of phagocytic cells, a population of >2n phagocytic cells, a population of circulating vesicles, or a population of circulating diseased cells in place of the population of phagocytic cells, the population of >2n phagocytic cells, the population of circulating vesicles, or the population of circulating diseased cells, respectively. In some embodiments, the combination sample may comprise an analyte isolated from a cell-free bodily fluid, or a portion or fraction of an analyte isolated from a cell-free bodily fluid, rather than the cell-free bodily fluid itself. A control profile may be from phagocytic cells, phagocytic cells having a DNA content of 2n (=2n phagocytic cells), non-phagocytic cells, or control cells that are substantially free of cells affected by the disease or condition, taken from the same individual. Alternatively, the control profile may be a profile from a repository of markers of a disease or condition. In the context of this invention, "control cells that are substantially free of cells affected by the disease or condition" refers to a population of cells, as compared to circulating diseased cells, comprise significantly

fewer amounts of cells affected by the disease or condition. In some embodiments, "control cells that are substantially free of cells affected by the disease or condition" are cells that are at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 100% free of cells affected by the disease or condition. In some embodiments, control cells that can be used in the methods of this invention are substantially free of fetal material (e.g., nucleic acids, proteins, and any analyte described herein), such as control cells that are at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 100% free of fetal material.

[0032] This invention also provides methods for assessing the risk of developing a disease or condition, prognosing said disease, monitoring said disease progression or regression, assessing the efficacy of a treatment, or identifying a compound capable of ameliorating or treating said disease or condition.

[0033] In some embodiments, one or more components of whole blood that are substantially free of markers of a disease or condition may be isolated from a whole blood sample. In these embodiments, the one or more components of whole blood that are substantially free of markers may be used as a control sample (e.g., to determine a control profile), and the remaining portion of the whole blood sample may be used as an analytical sample (e.g., to determine an analytical profile). For example, a population of non-phagocytic cells may be isolated from a whole blood sample and used to determine a control profile, while the remaining portion of the whole blood sample is used to determine an analytical profile.

[0034] The methods of this invention can be used together with any known diagnostic methods, such as physical inspection, visual inspection, biopsy, scanning, histology, radiology, imaging, ultrasound, use of a commercial kit, genetic testing, immunological testing, analysis of bodily fluids, or monitoring neural activity.

[0035] Phagocytic cells that can be used in the methods of this invention include all types of cells that are capable of ingesting various types of substances (e.g., apoptotic cells, infectious agents, dead cells, viable cells, cell-free DNAs, cell-free RNAs, cell-free proteins). In some embodiments, the phagocytic cells are neutrophils, macrophages, monocytes, dendritic cells, foam cells, mast cells, eosinophils, or

keratinocytes. In some embodiments, the phagocytic cells can be a mixture of different types of phagocytic cells. In some embodiments, the phagocytic cells can be activated phagocytic cells., e.g., activated macrophages or neutrohils. In some embodiments, a phagocyte is a histiocyte, e.g., a Langerhans cell.

[0036] As used herein, ">2n phagocytic cells" refer to phagocytic cells that have a DNA content of greater than 2n, while "=2n phagocytic cells" refer to phagocytic cells that have a DNA content of 2n. According to the present invention, some phagocytic cells engulf live/dying/dead diseased cells (and sub-cellular fragments thereof) and/or cell-free disease-specific nucleic acids, proteins, carbohydrates and/or lipids present in bodily fluids. Such phagocytosis leads to the internalization of these disease markers into the phagocytic cell and, therefore, the DNA content of these phagocytic cells will become greater than 2n. By contrast, some phagocytic cells have not engulfed living/dying/dead diseased cells or fragments and/or cell-free disease-specific nucleic acids, proteins, lipids, and/or carbohydrates present in bodily fluids. The DNA contents of this group of phagocytic cells remain 2n. In some embodiments, the disease-specific markers (e.g., DNA with disease-specific mutations) can be expressed by the >2n phagocytic cells. For example, the mutated DNA of diseased cells is integrated into the normal DNA of the >2n phagocytic cells. The subsequent transcription of the "integrated" DNA of the >2n phagocytic cells into RNA and the translation of the latter into proteins produces a phenotype different from the phagocytic cells that have not phagocytosed the diseased cells (i.e., the =2n phagocytic cells). In other embodiments, the internalized disease-specific markers are not expressed by the >2n phagocytic cells. The markers may be translocated onto the membranes of the >2n phagocytic cells, or secreted out by the >2n phagocytic cells.

[0037] Circulating diseased cells that can be used in the methods of this invention include all types of circulating cells that may be affected by a disease or condition or infected by an infectious agent. A circulating cell refers to a cell present in the bodily fluid. A circulating cell may not necessarily circulate throughout the entire body or in the circulatory system. For example, a circulating cell may be present locally, such as in synovial fluid, or cerebrospinal fluid, or lymph fluid. A circulating diseased cell

may also be detached from a tissue or organ that has been affected by a disease or condition or infected by an infectious agent.

[0038] Cells that can be used as control cells in the methods of this invention includes all types of normal cells, or healthy cells, or cells that are substantially free of a disease or condition, or cells that are substantially free of an infectious agent. Control cells may be circulating cells or non-circulating cells (e.g., biopsied cells) that are representative of a normal or non-diseased state to which measurements on circulating diseased cells are compared to determine whether one or more diseased-associated marker is present in different levels between the circulating diseased cells and the control cells. The nature of the control cell may be a matter of design choice for a particular assay and may be derived or determined from normal tissue of the patient him- or herself.

[0039] In some embodiments, the circulating diseased cells are blood cells, tumor cells, lymphoma cells, fetal cells, apoptotic cells, epithelia cells, endothelial cells, stem cells, progenitor cells, mesenchymal cells, osteoblast cells, osteocytes, hematopoietic stem cells, foam cells, adipose cells, transcervical cells, circulating cardiocytes, circulating fibrocytes, circulating cancer stem cells, circulating myocytes, circulating cells from kidney, circulating cells from gastrointestinal tract, circulating cells from lung, circulating cells from reproductive organs, circulating cells from central nervous system, circulating hepatic cells, circulating cells from spleen, circulating cells from thymus, circulating cells from thyroid , circulating cells from an endocrine gland, circulating cells from parathyroid, circulating cells from pituitary, circulating cells from adrenal gland, circulating cells from islets of Langerhans, circulating cells from pancreas, circulating cells from hypothalamus, circulating cells from prostate tissues, circulating cells from breast tissues, circulating cells from circulating retinal cells, circulating ophthalmic cells, circulating auditory cells, circulating epidermal cells, or circulating cells from the urinary tract. In other embodiments, the circulating diseased cells can be a mixture of different types of circulating diseased cells.

[0040] The circulating diseased cells that can be used in the methods of this invention may be affected by various diseases or conditions. Exemplary diseases or conditions are a cardiovascular disease or condition, a kidney-associated disease or condition, a prenatal or pregnancy-related disease or condition, a neurological or neuropsychiatric disease or condition, an autoimmune or immune-related disease or condition, a cancer, an infectious disease or condition, a mitochondrial disorder, a respiratory-gastrointestinal tract disease or condition, a reproductive disease or condition, an ophthalmic disease or condition, a musculo-skeletal disease or condition, or a dermal disease or condition.

[0041] The circulating diseased cells that can be used in the methods of this invention may be infected by an infectious agent, such as a virus, bacteria, fungus, parasite, protozoan, infectious protein or microorganism.

[0042] In some embodiments, a cell useful in the methods of the invention (e.g., a circulating diseased cell, control cell not affected by the disease or condition, phagocytic cell, >2n phagocytic cell, =2n phagocytic cell, or non-phagocytic cell) is enucleated. Cells may be enucleated, for example, by using physical removal (e.g., via microneedle, optical tweezer, or aspiration), chemical treatments, photoablation, or ultraviolet irradiation.

[0043] As used herein, a "circulating vesicle" refers to a membrane-bound vesicle of cellular origin. The circulating vesicle may not necessarily circulate throughout the entire body or in the circulatory system. For example, the circulating vesicle may be present locally, such as in synovial fluid, or cerebrospinal fluid, or lymph fluid. In some embodiments, the circulating vesicles are selected from the group consisting of circulating microvesicles, apoptotic bodies, micro-particles, membrane-bound vesicles, multivesicular bodies, nanovesicles, microparticles, and ARRDC-1 mediated microvesicles (ARMM). In further embodiments, the circulating microvesicles are exosomes or urinary exosomes.

[0044] In some embodiments, a combination sample comprises two or more different components selected from the group consisting of: a cell-free bodily fluid isolated from a subject, a population of phagocytic cells isolated from the subject, a

population of >2n phagocytic cells isolated from the subject, a population of circulating vesicles isolated from the subject, and a population of circulating diseased cells isolated from the subject. In some embodiments, a combination sample comprises an analyte isolated from a cell-free bodily fluid isolated from the subject, an analyte isolated from a population of phagocytic cells isolated from the subject, an analyte isolated from a population of >2n phagocytic cells isolated from the subject, an analyte isolated from a population of circulating vesicles isolated from the subject, and an analyte isolated from a population of circulating diseased cells isolated from the subject.

[0045] As used herein, a "profile" of a marker of a disease or condition can broadly refer to any information concerning the marker. This information can be either qualitative (e.g., presence or absence) or quantitative (e.g., levels, copy numbers, or dosages). In some embodiments, a profile of a marker can indicate the absence of this marker. The profile can be a nucleic acid (e.g., DNA or RNA) profile, a protein profile, a lipid profile, a carbohydrate profile, a metabolite profile, or a combination thereof. A "marker" as used herein generally refers to an analyte which is differentially detectable in phagocytes and is indicative of the presence of a disease or condition. An analyte is differentially detectable if it can be distinguished quantitatively or qualitatively in phagocytes.

[0046] The methods of this invention can be applied to various diseases or conditions. Exemplary diseases or conditions are a cardiovascular disease or condition, a kidney-associated disease or condition, a prenatal or pregnancy-related disease or condition, a neurological or neuropsychiatric disease or condition, an autoimmune or immune-related disease or condition, a cancer, an infectious disease or condition, a pediatric disease, disorder, or condition, a mitochondrial disorder, a respiratory-gastrointestinal tract disease or condition, a reproductive disease or condition, an ophthalmic disease or condition, a musculo-skeletal disease or condition, or a dermal disease or condition.

[0047] As used herein, the term "cardiovascular disease or condition" refers to any condition that affects systems of heart or blood vessels (arteries and veins). Examples

of cardiovascular diseases include, but are not limited to myocardial infarction, coronary artery disease, percutaneous transluminal coronary angioplasty (PTCA), coronary artery bypass surgery (CABG), restenosis, peripheral arterial disease, stroke, abdominal aorta aneurysm, intracranial aneurysm, large artery atherosclerotic stroke, cardiogenic stroke, an early onset myocardial infarction, heart failure, pulmonary embolism, acute coronary syndrome (ACS), angina, cardiac hypertrophy, arteriosclerosis, myocarditis, pancarditis, endocarditis, hypertension, congestive heart failure, atherosclerosis, cerebrovascular disease, declining cardiac health, ischemic heart disease, pericarditis, cardiogenic shock, alcoholic cardiomyopathy, congenital heart disease, ischemic cardiomyopathy, hypertensive cardiomyopathy, valvular cardiomyopathy, inflammatory cardiomyopathy, cardiomyopathy secondary to a systemic metabolic disease, dilated cardiomyopathy, hypertrophic cardiomyopathy, arrhythmogenic right ventricular cardiomyopathy, restrictive cardiomyopathy, noncompaction cardiomyopathy, valvular heart disease, hypertensive heart disease, myocardial ischemic attack, unstable angina, myocardial rupture, cardiogenic shock, embolism, deep vein thrombosis, arrhythmia, arrhythmogenic right ventricular cardiomyopathy, diabetic cardiomyopathy, mitral regurgitation, mitral valve prolapse, peripheral vascular disease, artery disease, carotid artery disease, deep vein thrombosis, venous diseases, cerebrovascular disease, arterial aneurysm, left ventricular hypertrophy, hypertensive renal disease, hypertensive retinal disease, vasculitis, left main disease, arterial vascular disease, venous vascular disease, thrombosis of the microcirculation, a transient cerebrovascular accident, limb ischemia, aneurysm, thrombosis, superficial venous thrombosis, and deep venous thrombosis.

[0048] As used herein, the term "kidney-associated disease or condition" refers to any disease or condition that affects kidney or renal system. Examples of kidney-associated disease include, but are not limited to, chronic kidney diseases, primary kidney diseases, non-diabetic kidney diseases, glomerulonephritis, interstitial nephritis, diabetic kidney diseases, diabetic nephropathy, glomerulosclerosis, rapid progressive glomerulonephritis, renal fibrosis, Alport syndrome, IDDM nephritis, mesangial proliferative glomerulonephritis, membrano proliferative glomerulonephritis, crescentic glomerulonephritis, renal insterstitial fibrosis, focal

segmental glomerulosclerosis, membranous nephropathy, minimal change disease, pauci-immune rapid progressive glomerulonephritis, IgA nephropathy, polycystic kidney disease, Dent's disease, nephrocytinosis, Heymann nephritis, autosomal dominant (adult) polycystic kidney disease, autosomal recessive (childhood) polycystic kidney disease, acute kidney injury, nephrotic syndrome, renal ischemia, podocyte diseases or disorders, proteinuria, glomerular diseases, membranous glomerulonephritis, focal segmental glomerulonephritis, pre-eclampsia, eclampsia, kidney lesions, collagen vascular diseases, benign orthostatic (postural) proteinuria, IgM nephropathy, membranous nephropathy, sarcoidosis, diabetes mellitus, kidney damage due to drugs, Fabry's disease, aminoaciduria, Fanconi syndrome, hypertensive nephrosclerosis, interstitial nephritis, Sickle cell disease, hemoglobinuria, myoglobinuria, Wegener's Granulomatosis, Glycogen Storage Disease Type 1, chronic kidney disease, chronic renal failure, low Glomerular Filtration Rate (GFR), nephroangiosclerosis, lupus nephritis, ANCA-positive pauci-immune crescentic glomerulonephritis, chronic allograft nephropathy, nephrotoxicity, renal toxicity, kidney necrosis, kidney damage, glomerular and tubular injury, kidney dysfunction, nephritic syndrome, acute renal failure, chronic renal failure, proximal tubal dysfunction, acute kidney transplant rejection, chronic kidney transplant refection, non IgA mesangioproliferative glomerulonephritis, postinfectious glomerulonephritis, vasculitides with renal involvement of any kind, any hereditary renal disease, any interstitial nephritis, renal transplant failure, kidney cancer, kidney disease associated with other conditions (e.g., hypertension, diabetes, and autoimmune disease), Dent's disease, nephrocytinosis, Heymann nephritis, a primary kidney disease, a collapsing glomerulopathy, a dense deposit disease, a cryoglobulinemia-associated glomerulonephritis, an Henoch-Schónlein disease, a postinfectious glomerulonephritis, a bacterial endocarditis, a microscopic polyangitis, a Churg-Strauss syndrome, an anti-GBM-antibidy mediated glomerulonephritis, amyloidosis, a monoclonal immunoglobulin deposition disease, a fibrillary glomerulonephritis, an immunotactoid glomerulopathy, ischemic tubular injury, a medication-induced tubulo-interstitial nephritis, a toxic tubulo-interstitial nephritis, an infectious tubulo-interstitial nephritis, a bacterial pyelonephritis, a viral infectious tubulo-interstitial nephritis which results from a polyomavirus infection or an HIV infection, a

metabolic-induced tubulo-interstitial disease, a mixed connective disease, a cast nephropathy, a crystal nephropathy which may results from urate or oxalate or drug-induced crystal deposition, an acute cellular tubulo-interstitial allograft rejection, a tumoral infiltrative disease which results from a lymphoma or a post-transplant lymphoproliferative disease, an obstructive disease of the kidney, vascular disease, a thrombotic microangiopathy, a nephroangiosclerosis, an atheroembolic disease, a mixed connective tissue disease, a polyarteritis nodosa, a calcineurin-inhibitor induced- vascular disease, an acute cellular vascular allograft rejection, an acute humoral allograft rejection, early renal function decline (ERFD), end stage renal disease (ESRD), renal vein thrombosis, acute tubular necrosis, acute interstitial nephritis, established chronic kidney disease, renal artery stenosis, ischemic nephropathy, uremia, drug and toxin-induced chronic tubulointerstitial nephritis, reflux nephropathy, kidney stones, Goodpasture's syndrome, and hydronephrosis.

[0049] As used herein, the term "prenatal or pregnancy-related disease or condition" refers to any disease, disorder, or condition affecting a pregnant woman, embryo, or fetus. Prenatal or pregancy-related conditions can also refer to any disease, disorder, or condition that is associated with or arises, either directly or indirectly, as a result of pregnancy. These diseases or conditions can include any and all birth defects, congenital conditions, or hereditary diseases or conditions. Examples of prenatal or pregnancy-related diseases include, but are not limited to, Rhesus disease, hemolytic disease of the newborn, beta-thalassemia, sex determination, determination of pregnancy, a hereditary Mendelian genetic disorder, chromosomal aberrations, a fetal chromosomal aneuploidy, fetal chromosomal trisomy, fetal chromosomal monosomy, trisomy 8, trisomy 13 (Patau Syndrom), trisomy 16, trisomy 18 (Edwards syndrome), trisomy 21 (Down syndrome), X-chromosome linked disorders, trisomy X (XXX syndrome), monosomy X (Turner syndrome), XXY syndrome, XYY syndrome, XYY syndrome, XXXY syndrome, XXYY syndrome, XYYY syndrome, XXXXX syndrome, XXXXY syndrome, XXXYY syndrome, XXYYY syndrome, Fragile X Syndrome, fetal growth restriction, cystic fibrosis, a hemoglobinopathy, fetal death, fetal alcohol syndrome, sickle cell anemia, hemophilia, Klinefelter syndrome, dup(17)(p11.2p1.2) syndrome, endometriosis, Pelizaeus-Merzbacher disease, dup(22)(q11.2q11.2) syndrome, cat eye syndrome, cri-du-chat syndrome, Wolf-

Hirschhorn syndrome, Williams-Beuren syndrome, Charcot-Marie-Tooth disease, neuropathy with liability to pressure palsies, Smith-Magenis syndrome, neurofibromatosis, Alagille syndrome, Velocardiofacial syndrome, DiGeorge syndrome, steroid sulfatase deficiency, Prader-Willi syndrome, Kallmann syndrome, microphthalmia with linear skin defects, adrenal hypoplasia, glycerol kinase deficiency, Pelizaeus-Merzbacher disease, testis-determining factor on Y, azospermia (factor a), azospermia (factor b), azospermia (factor c), 1p36 deletion, phenylketonuria, Tay-Sachs disease, adrenal hyperplasia, Fanconi anemia, spinal muscular atrophy, Duchenne's muscular dystrophy, Huntington's disease, myotonic dystrophy, Robertsonian translocation, Angelman syndrome, tuberous sclerosis, ataxia telangieltasia, open spina bifida, neural tube defects, ventral wall defects, small-for-gestational-age, congenital cytomegalovirus, achondroplasia, Marfan's syndrome, congenital hypothyroidism, congenital toxoplasmosis, biotinidase deficiency, galactosemia, maple syrup urine disease, homocystinuria, medium-chain acyl Co-A dehydrogenase deficiency, structural birth defects, heart defects, abnormal limbs, club foot, anencephaly, arhinencephaly/holoprosencephaly, hydrocephaly, anophthalmos/microphthalmos, anotia/microtia, transposition of great vessels, tetralogy of Fallot, hypoplastic left heart syndrome, coarctation of aorta, cleft palate without cleft lip, cleft lip with or without cleft palate, oesophageal atresia/stenosis with or without fistula, small intestine atresia/stenosis, anorectal atresia/stenosis, hypospadias, indeterminate sex, renal agenesis, cystic kidney, preaxial polydactyly, limb reduction defects, diaphragmatic hernia, blindness, cataracts, visual problems, hearing loss, deafness, X-linked adrenoleukodystrophy, Rett syndrome, lysosomal disorders, cerebral palsy, autism, aglossia, albinism, ocular albinism, oculocutaneous albinism, gestational diabetes, Arnold-Chiari malformation, CHARGE syndrome, congenital diaphragmatic hernia, brachydactlia, aniridia, cleft foot and hand, heterochromia, Dwarnian ear, Ehlers Danlos syndrome, epidermolysis bullosa, Gorham's disease, Hashimoto's syndrome, hydrops fetalis, hypotonia, Klippel-Feil syndrome, muscular dystrophy, osteogenesis imperfecta, progeria, Smith Lemli Opitz symdrom, chromatelopsia, X-linked lymphoproliferative disease, omphalocele, gastroschisis, pre-eclampsia, eclampsia, pre-term labor, premature birth, miscarriage,

delayed intrauterine growth, ectopic pregnancy, hyperemesis gravidarum, morning sickness, or likelihood for successful induction of labor.

[0050] As used herein, the term "a neurological or neuropsychiatric disease or condition" refers to any disease or condition that affects nervous systems. Examples of neurological or neuropsychiatric diseases or conditions include, but are not limited to, head trauma, stroke, stroke, ischemic stroke, hemorrhagic stroke, subarachnoid hemorrhage, intra cranial hemorrhage, transient ischemic attack, vascular dementia, corticobasal ganglionic degeneration, encephalitis, epilepsy, Landau-Kleffner syndrome, hydrocephalus, pseudotumor cerebri, thalamic diseases, meningitis, myelitis, movement disorders, essential tremor, spinal cord diseases, syringomyelia, Alzheimer's disease (early onset), Alzheimer's disease (late onset), multi-infarct dementia, Pick's disease, Huntingdon's disease, Parkinson's disease, Parkinson syndromes, dementia, frontotemporal dementia, corticobasal degeneration, multiple system atrophy, progressive supranuclear palsy, Lewy body disease, Creutzfeldt-Jakob disease, Dandy-Walker syndrome, Friedreich ataxia, Machado-Joseph disease, migraine, schizophrenia, mood disorders and depression. dementia with lewy bodies (DLB), frontotemporal dementia (FTD), various forms of vascular dementia (VD), subcortical vascular dementia (Binswanger's disease), autism, developmental retardations, motor neuron diseases, amyotrophic lateral sclerosis (ALS), neuronal or brain damage, hypoxia of the brain, cerebral palsy (CP), memory disorders, movement disorders, corticalbasal ganglionic degeneration, forms of multiple system atrophy, stroke-related disorders, cerebrovascular accidents, post-irradiation encephalopathy with seizures, vascular Parkinsonism, thalamic cerebrovascular accidents, chronic inflammatory demyelinating polyneuropathy, alcohol related dementia, semantic dementia, ataxia, atypical Parkinsonism, dystonia, progressive supranuclear palsy, essential tremor, mild cognitive impairment, amyotrophic lateral sclerosis, multiple sclerosis, neuropathies, Pick's disease, congophilic amyloid angiopathy, Creutzfeldt-Jakob Disease, AIDS dementia complex, depression, anxiety disorder, phobia, Bell's Palsy, epilepsy, encephalitis, neuromuscular disorders, neurooncological disorders, brain tumors, neurovascular disorders, neuroimmunological disorders, neurootological disease, neurotrauma including spinal cord injury, pain including neuropathic pain, pediatric neurological and

neuropsychiatric disorders, sleep disorders, Tourette syndrome, corticalbasal ganglionic degeneration, alcohol related dementia, semantic dementia, Alzheimer's disease combined with multi-infarct dementia, Alzheimer's disease combined with Lewy body dementia, Parkinson's disease combined with Lewy body dementia, Alzheimer's and Parkinson's disease combined with Lewy body dementia, frontotemporal dementia combined with chronic inflammatory demyelinating polyneuropathy, attention deficit hyperactivity disorder, schizophrenia, obsessive-compulsive disorder, mental retardation, autistic spectrum disorders, opsoclonus-myoclonus syndrome (OMS) seizures, articulation disorder, learning disabilities (i.e., reading or arithmetic), verbal or performance aptitude deficits, attention deficit disorder, amyloid diseases, prion diseases, Tauopathies, Alpha-Synucleinopathies, addictive states such as those caused by at least one of: cocaine, nicotine, alcohol, food, ecstasy, kat, caffeine, opium, heroin, marijuana, amphetamine, methamphetamine or gambling, and Fabry's disease.

[0051] As used herein, the term "an autoimmune or immune-related disease or condition" refers to any disease or condition that affects the function of immune systems. Examples of autoimmune or immune-related diseases or conditions include, but are not limited to, antiphospholipid syndrome, systemic lupus erythematosus, rheumatoid arthritis, autoimmune vasculitis, celiac disease, autoimmune thyroiditis, post-transfusion immunization, maternal-fetal incompatibility, transfusion reactions, immunological deficiency such IgA deficiency, common variable immunodeficiency, drug-induced lupus, diabetes mellitus, Type I diabetes, Type II diabetes, juvenile onset diabetes, juvenile rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, immunodeficiency, allergies, asthma, psoriasis, atopic dermatitis, allergic contact dermatitis, chronic skin diseases, amyotrophic lateral sclerosis, chemotherapy-induced injury, graft-vs-host diseases, bone marrow transplant rejection, Ankylosing spondylitis, atopic eczema, Pemphigus, Behcet's disease, chronic fatigue syndrome fibromyalgia, chemotherapy-induced injury, myasthenia gravis, glomerulonephritis, allergic retinitis, systemic sclerosis, subacute cutaneous lupus erythematosus, cutaneous lupus erythematosus including chilblain lupus erythematosus, Sjogren's syndrome, autoimmune nephritis, autoimmune vasculitis, autoimmune hepatitis, autoimmune carditis, autoimmune encephalitis, autoimmune mediated hematological

diseases, lc-SSc (limited cutaneous form of scleroderma), dc-SSc (diffused cutaneous form of scleroderma), autoimmune thyroiditis (AT), Grave's disease (GD), myasthenia gravis, multiple sclerosis (MS), ankylosing spondylitis. transplant rejection, immune aging, rheumatic/autoimmune diseases, mixed connective tissue disease, spondyloarthropathy, psoriasis, psoriatic arthritis, myositis, scleroderma, dermatomyositis, autoimmune vasculitis, mixed connective tissue disease, idiopathic thrombocytopenic purpura, Crohn's disease, human adjuvant disease, osteoarthritis, juvenile chronic arthritis, a spondyloarthropathy, an idiopathic inflammatory myopathy, systemic vasculitis, sarcoidosis, autoimmune hemolytic anemia, autoimmune thrombocytopenia, thyroiditis, immune-mediated renal disease, a demyelinating disease of the central or peripheral nervous system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome, a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary disease, infectious or autoimmune chronic active hepatitis, primary biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis, inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's disease, an autoimmune or immune-mediated skin disease, a bullous skin disease, erythema multiforme, allergic rhinitis, atopic dermatitis, food hypersensitivity, urticaria, an immunologic disease of the lung, eosinophilic pneumonias, idiopathic pulmonary fibrosis, hypersensitivity pneumonitis, a transplantation associated disease, graft rejection or graft-versus-host-disease, psoriatic arthritis, psoriasis, dermatitis, polymyositis/dermatomyositis, toxic epidermal necrolysis, systemic scleroderma and sclerosis, responses associated with inflammatory bowel disease, Crohn's disease, ulcerative colitis, respiratory distress syndrome, adult respiratory distress syndrome (ARDS), meningitis, encephalitis, uveitis, colitis, glomerulonephritis, allergic conditions, eczema, asthma, conditions involving infiltration of T cells and chronic inflammatory responses, atherosclerosis, autoimmune myocarditis, leukocyte adhesion deficiency, allergic encephalomyelitis, immune responses associated with acute and delayed hypersensitivity mediated by cytokines and T-lymphocytes, tuberculosis, sarcoidosis, granulomatosis including Wegener's granulomatosis, agranulocytosis, vasculitis (including ANCA), aplastic anemia, Diamond Blackfan anemia, immune hemolytic anemia including autoimmune hemolytic anemia (AIHA), pernicious anemia, pure red cell aplasia (PRCA), Factor VIII deficiency, hemophilia

A, autoimmune neutropenia, pancytopenia, leukopenia, diseases involving leukocyte diapedesis, central nervous system (CNS) inflammatory disorders, multiple organ injury syndrome, mysathenia gravis, antigen-antibody complex mediated diseases, anti-glomerular basement membrane disease, anti-phospholipid antibody syndrome, allergic neuritis, Bechet disease, Castleman's syndrome, Goodpasture's syndrome, Lambert-Eaton Myasthenic Syndrome, Reynaud's syndrome, Sjorgen's syndrome, Stevens-Johnson syndrome, pemphigoid bullous, pemphigus, autoimmune polyendocrinopathies, Reiter's disease, stiff-man syndrome, giant cell arteritis, immune complex nephritis, IgA nephropathy, IgM polyneuropathies or IgM mediated neuropathy, idiopathic thrombocytopenic purpura (ITP), thrombotic throbocytopenic purpura (TTP), autoimmune thrombocytopenia, autoimmune disease of the testis and ovary including autoimmune orchitis and oophoritis, primary hypothyroidism, autoimmune endocrine diseases including autoimmune thyroiditis, chronic thyroiditis (Hashimoto's Thyroiditis), subacute thyroiditis, idiopathic hypothyroidism, Addison's disease, Grave's disease, autoimmune polyglandular syndromes (or polyglandular endocrinopathy syndromes), Sheehan's syndrome, autoimmune hepatitis, lymphoid interstitial pneumonitis (HIV), bronchiolitis obliterans (non-transplant) vs NSIP, Guillain-Barre' Syndrome, large vessel vasculitis (including polymyalgia rheumatica and giant cell (Takayasu's) arteritis), medium vessel vasculitis (including Kawasaki's disease and polyarteritis nodosa), ankylosing spondylitis, Berger's disease (IgA nephropathy), rapidly progressive glomerulonephritis, primary biliary cirrhosis, Celiac sprue (gluten enteropathy), cryoglobulinemia, and amyotrophic lateral sclerosis (ALS).

[0052] As used herein, the term "cancer" refers to various types of malignant neoplasms, most of which can invade surrounding tissues, and may metastasize to different sites (see, for example, PDR Medical Dictionary, 1st edition (1995), incorporated herein by reference in its entirety for all purposes). The terms "neoplasm" and "tumor" refer to an abnormal tissue that grows by cellular proliferation more rapidly than normal and continues to grow after the stimuli that initiated proliferation is removed. Such abnormal tissue shows partial or complete lack of structural organization and functional coordination with the normal tissue which may be either benign (i.e., benign tumor) or malignant (i.e., malignant tumor).

Examples of general categories of cancer include, but are not limited to, carcinomas (i.e., malignant tumors derived from epithelial cells such as, for example, common forms of breast, prostate, lung and colon cancer), sarcomas (i.e., malignant tumors derived from connective tissue or mesenchymal cells), lymphomas (i.e., malignancies derived from hematopoietic cells), leukemias (i.e., malignancies derived from hematopoietic cells), germ cell tumors (i.e., tumors derived from totipotent cells. In adults most often found in the testicle or ovary; in fetuses, babies and young children, most often found on the body midline, particularly at the tip of the tailbone), blastic tumors (i.e., a typically malignant tumor which resembles an immature or embryonic tissue) and the like. Examples of the types of neoplasms intended to be encompassed by the present invention include but are not limited to those neoplasms associated with cancers of neural tissue, blood forming tissue, breast, skin, bone, prostate, ovaries, uterus, cervix, liver, lung, brain, larynx, gallbladder, pancreas, rectum, parathyroid, thyroid, adrenal gland, immune system, head and neck, colon, stomach, bronchi, and/or kidneys.

[0053] As used herein, the term "an infectious disease or condition" refers to any disease or condition that results from an infectious agent. Infectious agents include, but are not limited to bacteria, viruses, fungi, protozoa, infectious proteins, parasitic microbes, and other parasites. Examples of infectious diseases or conditions include, but are not limited to, bacterial infections, viral infections, fungal infections, protozoan infections, parasitic infections, hepatitis (e.g., hepatitis A, B, C, D, and E), herpes, influenza, human papillomavirus (HPV) infection, AIDS, anthrax, pneumonia (bacterial or viral), cellulitis, human parainfluenza, the common cold, Legionnaires' disease (Legionellosis), cholera, Creutzfeldt-Jakob disease (CJD), variant Creutzfeldt-Jakob disease (vCJD), fatal familial insomnia (FFI), Gerstmann-Sträussler-Scheinker (GSS) syndrome, Chlamydia, chicken pox, ebola hemorrhagic fever, Dengue fever, giardiasis, Lyme disease, malaria, measles, mumps, rubella, pertussis, gonorrhea, staphylococcal infection, streptococcal infection, pneumococcal infection, rabies, helicobacter pylori infection, respiratory syncitial virus infection, Rocky Mountain spotted fever, SARS, sepsis, tuberculosis, and West Nile fever.

[0054] Viruses include, but are not limited to, DNA or RNA animal viruses. As used herein, RNA viruses include, but are not limited to, virus families such as Picornaviridae (e.g., polioviruses), Reoviridae (e.g., rotaviruses), Togaviridae (e.g., encephalitis viruses, yellow fever virus, rubella virus), Orthomyxoviridae (e.g., influenza viruses), Paramyxoviridae (e.g., respiratory syncytial virus, measles virus, mumps virus, parainfluenza virus), Rhabdoviridae (e.g., rabies virus), Coronaviridae, Bunyaviridae, Flaviviridae, Filoviridae, Arenaviridae, Bunyaviridae and Retroviridae (e.g., human T cell lymphotropic viruses (HTLV), human immunodeficiency viruses (HIV)). As used herein, DNA viruses include, but are not limited to, virus families such as Papovaviridae (e.g., papilloma viruses), Adenoviridae (e.g., adenovirus), Herpesviridae (e.g., herpes simplex viruses), and Poxviridae (e.g., variola viruses).

[0055] Bacteria include, but are not limited to, gram positive bacteria, gram negative bacteria, acid-fast bacteria and the like. Gram positive bacteria include, but are not limited to, Actinomedurae, Actinomyces israelii, Bacillus anthracis, Bacillus cereus, Clostridium botulinum, Clostridium difficile, Clostridium perfringens, Clostridium tetani, Corynebacterium, Enterococcus faecalis, Listeria monocytogenes, Nocardia, Propionibacterium acnes, Staphylococcus aureus, Staphylococcus epiderm, Streptococcus mutans, Streptococcus pneumoniae and the like. Gram negative bacteria include, but are not limited to, Afipia felis, Bacteriodes, Bartonella bacilliformis, Bortadella pertussis, Borrelia burgdorferi, Borrelia recurrentis, Brucella, Calymmatobacterium granulomatis, Campylobacter, Escherichia coli, Francisella tularensis, Gardnerella vaginalis, Haemophilius aegyptius, Haemophilius ducreyi, Haemophilius influenziae, Heliobacter pylori, Legionella pneumophila, Leptospira interrogans, Neisseria meningitidia, Porphyromonas gingivalis, Providencia sturti, Pseudomonas aeruginosa, Salmonella enteridis, Salmonella typhi, Serratia marcescens, Shigella boydii, Streptobacillus moniliformis, Streptococcus pyogenes, Treponema pallidum, Vibrio cholerae, Yersinia enterocolitica, Yersinia pestis and the like. As used herein, acid-fast bacteria include, but are not limited to, Myobacterium avium, Myobacterium leprae, Myobacterium tuberculosis and the like. Other bacteria not falling into the other three categories include, but are not limited to, Bartonella henseiae, Chlamydia psittaci, Chlamydia trachomatis, Coxiella burnetii, Mycoplasma pneumoniae, Rickettsia akari, Rickettsia prowazekii, Rickettsia rickettsii, Rickettsia

tsutsugamushi, Rickettsia typhi, Ureaplasma urealyticum, Diplococcus pneumoniae, Ehrlichia chafensis, Enterococcus faecium, Meningococci and the like.

[0056] As used herein, fungi include, but are not limited to, Aspergilli, Candidae, Candida albicans, Coccidioides immitis, Cryptococci, and combinations thereof.

[0057] As used herein, parasitic microbes include, but are not limited to, Balantidium coli, Cryptosporidium parvum, Cyclospora cayatanensis, Encephalitozoa, Entamoeba histolytica, Enterocytozoon bieneusi, Giardia lamblia, Leishmaniae, Plasmodii, Toxoplasma gondii, Trypanosomae, trapezoidal amoeba and the like. Other parasites include worms (e.g., helminthes), particularly parasitic worms including, but not limited to, Nematoda (roundworms, e.g., whipworms, hookworms, pinworms, ascarids, filarids and the like), and Cestoda (e.g., tapeworms).

[0058] As used herein, infectious proteins include prions (e.g., PrP^{Sc} forms, the CJD prion, the vCJD prion, the FFI prion, and the GSS prion).

[0059] As used herein, the term "a pediatric disease, disorder, or condition" refers to any disease or disorder that affects infants or children or that begins during development or childhood. Examples of pediatric diseases, disorders, and conditions include, but are not limited to, autism, Kawasaki's disease, congenital deafness, pediatric cancers, Type I diabetes, congenital heart defects, tetralogy of Fallot, Duchenne Muscular Dystrophy, osteogenesis importfect, Krabe disease, Pompe disease, Gaucher disease, Fabry disease, Wolff-Parkinson-White syndrome, Hirschsprung's disease, Crohn's disease, Eagle-Barrett Syndrome, cystic fibrosis, irritable bowel syndrome, and cerebral palsy. Examples of pediatric conditions also include genetic attributes of the developing fetus. For example, a pediatric conditions include, but are not limited to, intelligence, eye color, hair color, and muscle type.

[0060] As used herein, "treating" a disease or condition refers to taking steps to obtain beneficial or desired results, including clinical results. Beneficial or desired clinical results include, but are not limited to, alleviation or amelioration of one or more symptoms associated with diseases or conditions.

[0061] As used herein, "administering" or "administration of" a compound or an agent to a subject can be carried out using one of a variety of methods known to those skilled in the art. For example, a compound or an agent can be administered, intravenously, arterially, intradermally, intramuscularly, intraperitonealy, intravenously, subcutaneously, ocularly, sublingually, orally (by ingestion), intranasally (by inhalation), intraspinally, intracerebrally, and transdermally (by absorbtion, e.g., through a skin duct). A compound or agent can also appropriately be introduced by rechargeable or biodegradable polymeric devices or other devices, e.g., patches and pumps, or formulations, which provide for the extended, slow, or controlled release of the compound or agent. Administering can also be performed, for example, once, a plurality of times, and/or over one or more extended periods. In some aspects, the administration includes both direct administration, including self-administration, and indirect administration, including the act of prescribing a drug. For example, as used herein, a physician who instructs a patient to self-administer a drug, or to have the drug administered by another and/or who provides a patient with a prescription for a drug is administering the drug to the patient. In some embodiments, a compound or an agent is administered orally, e.g., to a subject by ingestion, or intravenously, e.g., to a subject by injection. In some embodiments, the orally administered compound or agent is in an extended release or slow release formulation, or administered using a device for such slow or extended release.

[0062] In certain embodiments, markers used in the methods of invention are up-regulated or activated in the combination sample compared to the phagocytic cells, =2n phagocytic cells, non-phagocytic cells, or control cells that are substantially free of cells affected by the disease or condition. Different diseases or conditions can be associated with either up-regulation (or activation) or down-regulation (or inhibition) of different markers. As used herein, "up-regulation or up-regulated" can refer to an increase in expression levels (e.g., gene expression or protein expression), gene copy numbers, gene dosages, and other qualitative or quantitative detectable state of the markers. Similarly, "down-regulation or down-regulated" can refer to an increase in expression levels, gene copy numbers, gene dosages, and other qualitative or quantitative detectable state of the markers. In certain embodiments, markers used in the methods of invention are down-regulated or inhibited in the combination sample

compared to the phagocytic cells, =2n phagocytic cells, non-phagocytic cells, or control cells that are substantially free of cells affected by the disease or condition. Different diseases or conditions can be associated with either up-regulation (or activation) or down-regulation (or inhibition) of different markers. As used herein, "activation or activated" can refer to an active state of the marker, e.g., a phosphorylation state, a DNA methylation state, or a DNA acetylation state. Similarly, "inhibition or inhibited" can refer to a repressed state or an inactivated state of the marker, e.g., a de-phosphorylation state, a ubiquitination state, a DNA de-methylation state.

[0063] In certain embodiments, methods of this invention can comprise extracting or enriching markers from cell-free bodily fluids. Any known extraction and enrichment methods can be used herein. In certain embodiments, methods of this invention also comprise at least one of the following steps before determination of various profiles: i) lysing the =2n phagocytic, non-phagocytic, or control cells; and ii) extracting cellular contents from the lysed cells. Any known cell lysis and extraction methods can be used herein. In certain embodiments, at least one or more markers of a disease or condition are present in the combination sample. In certain embodiments, there is no marker present in the cellular contents of the non-phagocytic, =2n phagocytic, or control cells.

[0064] In certain embodiments, methods of this invention also comprise at least one of the following steps before determination of various profiles: i) lysing the phagocytic, >2n phagocytic cells, or =2n phagocytic cells; and ii) extracting cellular contents from the lysed phagocytic, >2n phagocytic, or =2n phagocytic cells. In certain embodiments, the cellular contents of the phagocytic or >2n phagocytic cells comprise various types of materials that they have engulfed, such as viable diseased cells, dead diseased cells, apoptotic diseased cells, circulating tumor cells, infectious agents, fetal cells, trophoblasts, or fragments thereof. In certain embodiments, at least one or more markers of a disease or condition are present in the cellular contents of the phagocytic or >2n phagocytic cells.

[0065] In certain embodiments, methods of this invention also comprise at least one of the following steps before determination of various profiles: i) lysing the circulating vesicles; and ii) extracting the contents from the lysed circulating vesicles. In certain embodiments, the contents of the circulating vesicles comprise various types of materials, such as proteins and nucleic acids. In certain embodiments, at least one or more markers of a disease or condition are present in the contents of the circulating vesicles.

[0066] In certain embodiments, methods of this invention also comprise at least one of the following steps before determination of various profiles: i) lysing the circulating diseased cells; and ii) extracting the contents from the lysed circulating diseased cells. In certain embodiments, at least one or more markers of a disease or condition are present in the contents of the circulating diseased cells.

[0067] In certain embodiments, methods of this invention further comprise comparing the identified difference of the disease or condition-specific markers to a repository of at least one markers known in the art. Such comparison can further confirm the presence of the disease or condition. In some embodiments, the repository of the known markers can be obtained by data mining. The term "data mining", as used herein, refers to a process of finding new data patterns, relations, or correlations derived from the known data of the databases and of extracting practicable information in the future. Typically a computer-based system can be trained on data to perform the data mining, e.g., to classify the input data and then subsequently used with new input data to make decisions based on the training data. These systems include, but are not limited, expert systems, fuzzy logic, non-linear regression analysis, multivariate analysis, decision tree classifiers, and Bayesian belief networks.

[0068] In certain embodiments, the phagocytic cells, >2n phagocytic cells, circulating vesicles, circulating diseased cells, control cells, and =2n phagocytic cells are isolated from a bodily fluid sample, tissues, or cells. Exemplary bodily fluid samples can be whole blood, urine, stool, saliva, lymph fluid, cerebrospinal fluid, synovial fluid, cystic fluid, ascites, pleural effusion, fluid obtained from a pregnant

woman in the first trimester, fluid obtained from a pregnant woman in the second trimester, fluid obtained from a pregnant woman in the third trimester, maternal blood, amniotic fluid, chorionic villus sample, fluid from a preimplantation embryo, maternal urine, maternal saliva, placental sample, fetal blood, lavage and cervical vaginal fluid, interstitial fluid, buccal swab sample, sputum, bronchial lavage, Pap smear sample, or ocular fluid. In some embodiments, the phagocytic cells, >2n phagocytic cells, and =2n phagocytic cells are isolated from white blood cells. In certain embodiments, the >2n phagocytic cells and the =2n phagocytic cells are separated from a population of phagocytic cells.

[0069] In certain embodiments, the cell-free bodily fluid comes from a bodily fluid sample. Exemplary bodily fluid samples can be whole blood, urine, stool, saliva, lymph fluid, cerebrospinal fluid, synovial fluid, cystic fluid, ascites, pleural effusion, fluid obtained from a pregnant woman in the first trimester, fluid obtained from a pregnant woman in the second trimester, fluid obtained from a pregnant woman in the third trimester, maternal blood, amniotic fluid, chorionic villus sample, fluid from a preimplantation embryo, maternal urine, maternal saliva, placental sample, fetal blood, lavage and cervical vaginal fluid, interstitial fluid, buccal swab sample, sputum, bronchial lavage, Pap smear sample, or ocular fluid. In some embodiments, the cell-free bodily fluids are obtained by separating cells from the bodily fluid sample by methods known in the art, such as extraction, centrifugation, and filtration.

[0070] In some embodiments, a component for use in a combination sample may be obtained by removing cells from a bodily fluid. In some embodiments, a component for use in a combination sample may be obtained by destroying (e.g., lysing) cells in a bodily fluid. These embodiments may be taken in combination, for example, by removing some populations of cells and destroying other populations of cells. In some embodiments, a sample of whole blood may be used to create a combination sample, for example, by removing red blood cells, serum, and T cells, and using the remainder as the combination sample.

[0071] In certain embodiments, tissue or fluid samples including cells having a DNA content of 2n are obtained post separation (e.g., via centrifugation) of non-

cellular fraction of fluids obtained by puncture of a vein or artery followed by the withdrawal of blood, tissue biopsies, bronchoalveolar lavage, nasal lavage, eye lavage, peritoneal cavity lavage, vaginal lavage, bladder lavage, rectal lavage, fine needle aspiration of spinal fluid, synovial fluid aspiration, and the like. Cell free bodily fluids are obtained post separation (e.g., via centrifugation) of cellular fraction of fluids obtained by puncture of a vein or artery followed by the withdrawal of blood, tissue biopsies, bronchoalveolar lavage, nasal lavage, eye lavage, peritoneal cavity lavage, vaginal lavage, bladder lavage, rectal lavage, fine needle aspiration of spinal fluid, synovial fluid aspiration, and the like.

[0072] In the methods of this invention, cell separation/isolation/purification methods are used to isolate populations of cells from bodily fluid sample, cells, or tissues of a subject. A skilled worker can use any known cell separation/isolation/purification techniques to isolate phagocytic cells, >2n phagocytic cells, diseased cells, control cells, and =2n phagocytic cells from a bodily fluid, or to separate >2n phagocytic cells from =2n phagocytic cells. Exemplary techniques include, but are not limited to, using antibodies, flow cytometry, fluorescence activated cell sorting, filtration, gradient-based centrifugation, elution, microfluidics, magnetic separation technique, fluorescent-magnetic separation technique, nanostructure, quantum dots, high throughput microscope-based platform, or a combination thereof.

[0073] In the methods of this invention, cell separation/isolation/purification methods are used to isolate populations of circulating diseased cells from bodily fluid sample, cells, or tissues of a subject. Circulating diseased cells may be rare or in low quantity in a bodily fluid. Therefore, enrichment techniques (e.g., magnetic enrichment) may be used to enrich circulating diseased cells before the isolation. A skilled worker can use any known cell separation/isolation/purification techniques to isolate circulating diseased cells from a bodily fluid. Exemplary techniques include, but are not limited to, using antibodies, flow cytometry, fluorescence activated cell sorting, filtration, gradient-based centrifugation, elution, microfluidics, magnetic separation technique, fluorescent-magnetic separation technique, nanostructure, quantum dots, high throughput microscope-based platform, micro-fluidic technique,

fiber-optic array–scanning technique, laser-scanning cytometry technique, multiphoton intravital flow cytometry, photoacoustic flowmetry, nanoparticles targeting cell surface antigens, staining circulating diseased cells with detectable secreted products, or a combination thereof. Circulating diseased cells may have different physical properties compared to normal circulating cells, such as difference in size, density, charge, migratory properties, and some properties of specific cell types (e.g., melanocytic granules in circulating melanoma cells). A skilled worker can use any known cell separation/isolation/purification techniques based on such different properties to isolate circulating diseased cells. For example, differences in buoyant density may be used to separate circulating diseased cells (e.g., circulating tumor cells) from normal blood cells through gradient centrifugation. Filtration-based approaches may be used isolate circulating diseased cells (e.g., circulating tumor cells) based on their increased sizes compared to normal circulating cells. Antibody-based isolation approaches may be used to capture circulating diseased cells, which express epithelia cell surface markers that are absent from normal circulating blood cells. For example, conjugation of antibodies against epithelial cell adhesion molecule (EpCAM) to magnetic beads, followed by purification of captured cells through a magnetic field, may be used to enrich circulating tumor cells from the blood of patients with cancers of the breast, prostate, and colon. In certain embodiments, circulating diseased cells (e.g., transcervical cells) may be collected by the RareCellect™ device (Genetic Technologies) or similar devices.

[0074] In certain embodiments, the phagocytic cells, >2n phagocytic cells, diseased cells, control cells, and =2n phagocytic cells are isolated by using a product secreted by the cells. In certain embodiments, the phagocytic cells, >2n phagocytic cells, diseased cells, control cells, and =2n phagocytic cells are isolated by using a cell surface target (e.g., receptor protein) on the surface of the cells. In some embodiments, the cell surface target is a protein that has been engulfed by >2n phagocytic cells. In some embodiments, the cell surface target is expressed by cells on their plasma membranes. In some embodiments, the cell surface target is an exogenous protein that is translocated on the plasma membranes, but not expressed by the cells (e.g., the >2n phagocytic cells). In some embodiments, the cell surface target is a marker of the disease or condition to be detected.