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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
09 September 2022 (09.09.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/187695 A1

(51) International Patent Classification:
A61K 48/00 (2006.01) *C12Q 1/68* (2018.01)
C12Q 1/00 (2006.01) *G01N 33/00* (2006.01)

(21) International Application Number:
PCT/US2022/019010

(22) International Filing Date:
04 March 2022 (04.03.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/157,437 05 March 2021 (05.03.2021) US

(71) Applicant: **MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH** [US/US]; 200 First Street S.W., Rochester, Minnesota 55905 (US).

(72) Inventors: **TAYLOR, William R.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **MAHONEY, Douglas W.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **KISIEL, John B.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US).

(72) Inventor: **AHLQUIST (DECEASED), David A.** *(deceased)*.

(74) Agent: **GOETZ, Robert A.**; Casimir Jones S.C., 2275 Deming Way, Ste 310, Middleton, Wisconsin 53562 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))*
— *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))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: DETECTING CERVICAL CANCER

(57) Abstract: Provided herein is technology for cervical cancer screening. In particular, provided herein are methods, compositions, and related uses for detecting the presence or absence of cervical cancer, cervical pre-cancers (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia), and cervical cancer subtypes (e.g., cervical adenocarcinoma, squamous cell cervical cancer), or for discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers) from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample).

WO 2022/187695 A1

DETECTING CERVICAL CANCER

CROSS-REFERENCE TO RELATED APPLICATION

The present application claims priority to U.S. Provisional Patent Application No. 63/157,437, filed March 5, 2021 which is hereby incorporated by reference in its entirety.

FIELD OF INVENTION

Provided herein is technology for cervical cancer screening. In particular, provided herein are methods, compositions, and related uses for detecting the presence or absence of cervical cancer, cervical pre-cancers (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia), and cervical cancer subtypes (e.g., cervical adenocarcinoma, squamous cell cervical cancer), or for discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers) from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample).

BACKGROUND

Cervical cancer (CC) screening methods continue to evolve and are highly sensitive for the presence of cervical dysplasia and CC. However, the specificity of current molecular-based testing (i.e. high risk-HPV alone) is limited given the prevalence of high risk-HPV infections that are cleared without neoplastic transformation. As such, there is an urgent need for improved diagnostic tools for detecting cervical cancer from a single biological sample.

The present invention addresses this need.

SUMMARY

Provided herein is technology for cervical cancer screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence or absence of cervical cancer, cervical pre-cancers (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia), and cervical cancer subtypes (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and for discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers).

Indeed, as described in Examples I and II, experiments conducted during the course for identifying embodiments for the present invention identified a novel set of differentially methylated regions (DMRs) for discriminating cancer and pre-cancers of the cervix derived DNA from non-neoplastic control DNA, and from other types of gynecological cancers (e.g., endometrial, and ovarian cancers).

Such experiments list and describe 423 novel DNA methylation markers distinguishing cervical cancer (cervical cancer subtypes) and pre-cancers from benign cervical samples (see, Tables I-IV, VI-VIII, Example I), and for discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers). (see, Tables XI and XII, Example II).

From these 423 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing cervical cancer from benign cervical samples:

- Any of the markers recited in Table I (see, Example I);
- Any of the markers recited in Table III (see, Example I);
- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5 (Example I);
- C1orf114, MAX.chr6.58147682-58147771, ZNF773, NEUROG3, ASCL1, NID2, ZNF781, CRHR2, and MAX.chr9.36739811-36739868 (see, Table VI and Example I); and
- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18 (see, Table VIII, Example I).

From these 423 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing cervical adenocarcinoma from benign cervical samples:

- ABCB1, ARHGAP12, ASCL1, BARHL1, C1orf114, C2orf40, CACNA1C, CRHR2, HOPX_C, KCNQ5, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, PRDM12, SLC9A3, TMEM200C, TTYH1, ZNF382, ZNF773, and ZNF781 (see, Table VII, Example I).

From these 423 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing cervical squamous cell cancer from benign cervical samples:

- ABCB1, ARHGAP12, ASCL1, ATP10A, BARHL1, C1orf114, CACNA1C, CRHR2, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, TMEM200C, TTYH1, ZNF382, ZNF69, ZNF773, and ZNF781 (see, Table VII, Example I).

From these 423 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing cervix related in-situ adenocarcinoma (AIS) from benign cervical samples:

- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ZNF773, TTYH1, NEUROG3, ZNF781, MAX.chr9.36739811-36739868, CRHR2, and NID2 (see, Table VI and Example I); and
- ABCB1, ARHGAP12, ASCL1, BARHL1, C1orf114, C2orf40, CACNA1C, CRHR2, HOPX_C, KCNQ5, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, PRDM12, SLC9A3, TMEM200C, TTYH1, ZNF382, ZNF773, and ZNF781 (see, Table VII, Example I).

From these 423 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing in cervical intraepithelial neoplasia (CIN) from benign cervical samples:

- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ZNF773, TTYH1, NEUROG3, ZNF781, MAX.chr9.36739811-36739868, CRHR2, and NID2 (see, Table VI and Example I); and
- ABCB1, ARHGAP12, ASCL1, ATP10A, BARHL1, C1orf114, CACNA1C, CRHR2, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, TMEM200C, TTYH1, ZNF382, ZNF69, ZNF773, and ZNF781 (see, Table VII, Example I).

From these 423 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing cervical cancer from endometrial cancer and/or ovarian cancer:

- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18 (see, Table VIII, Example I);
- Any of the markers recited in Table X (see, Example I); and
- AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1 (see, Tables X and XI, Example II).

As described herein, the technology provides a number of methylated DNA markers (MDMs) and subsets thereof (e.g., sets of 2, 3, 4, 5, 6, 7, 8, 20, 50, 100, 150, 200, 300, 400, 423 markers) with high discrimination for cervical cancer, various types of cervical cancer (e.g., cervical adenocarcinoma, squamous cell cervical cancer), various types of cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia), and for discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers). Cervical intraepithelial neoplasia (CIN) can be characterized as CIN 1 which refers to abnormal cells affecting about one-third of the thickness of the cervical epithelium, CIN 2 which refers to abnormal cells affecting about two-thirds to two-thirds of the cervical epithelium, and CIN 3 which refers to abnormal cells affecting more than two-thirds of the cervical epithelium.

In certain embodiments, the present disclosure provides methods for characterizing a biological sample comprising measuring a methylation level of one or more methylated markers selected from Tables I, III, and X in the biological sample, wherein measuring a methylation level of one or more methylated markers comprises treating DNA from the biological sample with a reagent that modifies DNA in a methylation-specific manner.

In some embodiments, the biological sample is from a human subject. In some embodiments, the human subject has or is suspected of having cervical cancer, a cervical cancer subtype, and/or a cervical pre-cancer.

In some embodiments, the biological sample is selected from a tissue sample, a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a cerebrospinal fluid (CSF) sample, a saliva sample, a urine sample, and a stool sample. In some embodiments, the tissue sample is a cervical tissue sample. In some embodiments, the cervical tissue sample further comprises one or more of vaginal tissue, vaginal cells, endometrial tissue, endometrial cells, ovarian tissue, and ovarian cells. In some embodiments, the secretion sample is a cervical secretion sample. In some embodiments, the cervical secretion sample further comprises one or more of a vaginal secretion, an endometrial secretion, and an ovarian secretion. In some embodiments, the biological sample is collected with a collection device having an absorbing member capable of collecting the biological sample upon contact with a bodily region. In some embodiments, the absorbing member is a sponge having a shape and size suitable for insertion into a body orifice. In some embodiments, the collection device is a tampon (e.g., a standard tampon), a lavage that releases liquid into the vagina and re-collects fluid (e.g., a Pantarhei screener), a cervical brush (e.g., a brush inserted into the vagina and turned around to collect cells), a Fournier cervical self-sampling device (a tampon-like plastic wand), or a swab.

In some embodiments, the measured methylation level of the one or more methylation markers is compared to a methylation level of a corresponding one or more methylation markers in control samples without cervical cancer.

In some embodiments, the method further comprises determining that the individual has cervical cancer when the methylation level measured in the one or more methylation markers is higher than the methylation level measured in the respective control samples. In some embodiments wherein the method further comprises determining that the individual has

cervical cancer, the one or more methylated markers are selected from one of the following groups:

- the methylated markers recited in Tables I and/or III;
- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5 (Example I);
- C1orf114, MAX.chr6.58147682-58147771, ZNF773, NEUROG3, ASCL1, NID2, ZNF781, CRHR2, and MAX.chr9.36739811-36739868 (see, Table VI and Example I); and
- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18 (see, Table VIII, Example I).

In some embodiments, the method further comprises determining that the individual has a subtype of cervical cancer. In some embodiments, the subtype of cervical cancer is selected from cervical adenocarcinoma and squamous cell cervical cancer. In some embodiments, wherein the method further comprises determining that the individual has a subtype of cervical cancer, the one or more methylated markers are selected from one of the following groups:

- ABCB1, ARHGAP12, ASCL1, BARHL1, C1orf114, C2orf40, CACNA1C, CRHR2, HOPX_C, KCNQ5, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, PRDM12, SLC9A3, TMEM200C, TTYH1, ZNF382, ZNF773, and ZNF781 (see, Table VII, Example I).

In some embodiments, the method further comprises determining that the individual has a cervical pre-cancer. In some embodiments, the cervical pre-cancer is selected from cervix related in-situ adenocarcinoma and cervical intraepithelial neoplasia. In some embodiments, wherein the method further comprises determining that the individual has a

cervical pre-cancer, the one or more methylated markers are selected from one of the following groups:

- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ZNF773, TTYH1, NEUROG3, ZNF781, MAX.chr9.36739811-36739868, CRHR2, and NID2 (see, Table VI and Example I); and
- ABCB1, ARHGAP12, ASCL1, BARHL1, C1orf114, C2orf40, CACNA1C, CRHR2, HOPX_C, KCNQ5, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, PRDM12, SLC9A3, TMEM200C, TTYH1, ZNF382, ZNF773, and ZNF781 (see, Table VII, Example I).

In some embodiments, the measured methylation level of the one or more methylation markers is compared to a methylation level of a corresponding one or more methylation markers in endometrial cancer samples and/or ovarian cancer samples. In some embodiments, the method further comprises discriminating cervical cancer from endometrial cancer and/or ovarian cancer. In some embodiments wherein the method further comprises discriminating cervical cancer from endometrial cancer and/or ovarian cancer, the one or more methylated markers are selected from one of the following groups:

- the markers recited in Table X;
- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18 (see, Table VIII, Example I); and
- AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1 (see, Tables X and XI, Example II).

In some embodiments, the reagent that modifies DNA in a methylation-specific manner is a borane reducing agent. In some embodiments, the borane reducing agent is 2-picoline borane. In some embodiments, the reagent that modifies DNA in a methylation-specific manner comprises one or more of a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent. In some embodiments, the reagent that modifies DNA in a methylation-specific manner is a bisulfite reagent, and the treating produces bisulfite-treated DNA.

In some embodiments, the treated DNA is amplified with a set of primers specific for the one or more methylated markers. In some embodiments, the set of primers specific for the one or more methylated markers is selected from the group recited in Tables V and XII. In some embodiments, the set of primers specific for the one or more methylated markers is capable of binding an amplicon bound by a primer sequence for the specific methylated marker gene recited in Tables V and XII, wherein the amplicon bound by the primer sequence for the methylated marker gene recited in Tables V and XII is at least a portion of a genetic region for the methylated marker recited in Tables I, III, and X. In some embodiments, the set of primers specific for the one or more methylated markers is a set of primers that specifically binds at least a portion of a genetic region comprising chromosomal coordinates for a methylated marker recited in Tables I, III, and X.

In some embodiments, measuring a methylation level of one or more methylated markers comprises multiplex amplification. In some embodiments, measuring a methylation level of one or more methylated markers comprises using one or more methods selected from the group consisting of methylation-specific PCR, quantitative methylation-specific PCR, methylation-specific DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, flap endonuclease assay, PCR-flap assay, and bisulfite genomic sequencing PCR. In some embodiments, measuring a methylation level of one or more methylated markers comprises measuring methylation of a CpG site for the one or more methylation markers. In some embodiments, the CpG site is present in a coding region or a regulatory region. In some embodiments, the one or more methylated markers is described by the genomic coordinates shown in Tables I, III, and X.

In certain embodiments, the present disclosure provides methods for preparing a deoxyribonucleic acid (DNA) fraction from a biological sample useful for analyzing one or more genetic loci involved in one or more chromosomal aberrations, comprising:

(a) extracting genomic DNA from a biological sample;

(b) producing a fraction of the extracted genomic DNA by:

(i) treating the extracted genomic DNA with a reagent that modifies DNA in a methylation-specific manner;

(ii) amplifying the treated genomic DNA using separate primers specific for one or more methylation markers recited in Tables I, III, and X;

(c) analyzing one or more genetic loci in the produced fraction of the extracted genomic DNA by measuring a methylation level for each of the one or more methylation markers.

In some embodiments, the reagent that modifies DNA in a methylation-specific manner is a borane reducing agent. In some embodiments, the borane reducing agent is 2-picoline borane. In some embodiments, wherein the reagent that modifies DNA in a methylation-specific manner comprises one or more of a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent. In some embodiments, the reagent that modifies DNA in a methylation-specific manner is a bisulfite reagent, and the treating produces bisulfite-treated DNA.

In some embodiments, the set of primers specific for the one or more methylated markers is selected from the group recited in Tables V and XII. In some embodiments, the set of primers specific for the one or more methylated markers is capable of binding an amplicon bound by a primer sequence for the specific methylated marker gene recited in Tables V and XII, wherein the amplicon bound by the primer sequence for the methylated marker gene recited in Tables V and XII is at least a portion of a genetic region for the methylated marker recited in Tables I, III, and X. In some embodiments, the set of primers specific for the one or more methylated markers is a set of primers that specifically binds at least a portion of a genetic region comprising chromosomal coordinates for a methylated marker recited in Tables I, III, and X.

In some embodiments, measuring a methylation level of one or more methylated markers comprises multiplex amplification. In some embodiments, measuring a methylation level of one or more methylated markers comprises using one or more methods selected from the group consisting of methylation-specific PCR, quantitative methylation-specific PCR, methylation-specific DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, flap endonuclease assay, PCR-flap assay, and bisulfite genomic sequencing PCR. In some embodiments, measuring a methylation level of one or more methylated markers comprises measuring methylation of a CpG site for the one or more methylation markers. In some embodiments, the CpG site is present in a coding region or a regulatory region. In some embodiments, the one or more methylated markers is described by the genomic coordinates shown in Tables I, III, and X.

In some embodiments, the biological sample is selected from a tissue sample, a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g.,

cervical secretion, vaginal secretion), an organ secretion sample, a cerebrospinal fluid (CSF) sample, a saliva sample, a urine sample, and a stool sample. In some embodiments, the tissue sample is a cervical tissue sample. In some embodiments, the cervical tissue sample further comprises one or more of vaginal tissue, vaginal cells, endometrial tissue, endometrial cells, ovarian tissue, and ovarian cells. In some embodiments, the secretion sample is a cervical secretion sample. In some embodiments, the cervical secretion sample further comprises one or more of a vaginal secretion, an endometrial secretion, and an ovarian secretion. . In some embodiments, the biological sample is collected with a collection device having an absorbing member capable of collecting tissue and/or cells upon contact with a bodily region. In some embodiments, the absorbing member is a sponge having a shape and size suitable for insertion into a body orifice. In some embodiments, the collection device is a tampon (e.g., a standard tampon), a lavage that releases liquid into the vagina and re-collects fluid (e.g., a Pantarhei screener), a cervical brush (e.g., a brush that women insert into the vagina and is turned around to collect cells), a Fournier cervical self-sampling device (a tampon-like plastic wand), or a swab.

In some embodiments, the biological sample is from a human subject. In some embodiments, the human subject has or is suspected of having cervical cancer, a cervical cancer subtype, and/or a cervical pre-cancer.

In some embodiments, the one or more methylated markers are selected from one of the following groups:

- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5 (Example I);
- C1orf114, MAX.chr6.58147682-58147771, ZNF773, NEUROG3, ASCL1, NID2, ZNF781, CRHR2, and MAX.chr9.36739811-36739868 (see, Table VI and Example I);
- ABCB1, ARHGAP12, ASCL1, BARHL1, C1orf114, C2orf40, CACNA1C, CRHR2, HOPX_C, KCNQ5, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2,

NXPH1, PRDM12, SLC9A3, TMEM200C, TTYH1, ZNF382, ZNF773, and ZNF781 (see, Table VII, Example I);

- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18 (see, Table VIII, Example I); and
- AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1 (see, Tables X and XI, Example II).

In certain embodiments, the technology is related to assessing the presence of and methylation state of one or more of the MDMs described herein in a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample). These MDMs comprise one or more differentially methylated regions (DMR) as discussed herein, e.g., as provided in Tables I, III, and X. Methylation state is assessed in embodiments of the technology. As such, the technology provided herein is not restricted in the method by which a gene's methylation state is measured and thus the methylation state of a gene may be measured by any method know in the art.

Also provided herein are compositions and kits for practicing the methods. For example, in some embodiments, reagents (e.g., primers, probes) specific for one or more MDMs are provided alone or in sets (e.g., sets of primers pairs for amplifying a plurality of markers). Additional reagents for conducting a detection assay may also be provided (e.g., enzymes, buffers, positive and negative controls for conducting QuARTS, PCR, sequencing, bisulfite, Ten-Eleven Translocation (TET) enzyme (e.g., human TET1, human TET2, human TET3, murine TET1, murine TET2, murine TET3, Naegleria TET (NgTET), Coprinopsis cinerea (CcTET)), or a variant thereof), organic borane, or other assays). In some embodiments, the kits contain a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, Ten-Eleven Translocation (TET) enzyme (e.g., human TET1, human TET2, human TET3, murine TET1, murine TET2, murine TET3, Naegleria TET (NgTET), Coprinopsis cinerea (CcTET)), or a variant thereof), organic

borane). In some embodiments, the kits containing one or more reagents necessary, sufficient, or useful for conducting a method are provided. Also provided are reactions mixtures containing the reagents. Further provided are master mix reagent sets containing a plurality of reagents that may be added to each other and/or to a test sample to complete a reaction mixture. In some embodiments, the kit comprises a control nucleic acid comprising one or more sequences from DMR 1–423 (from Tables I, III, and X) and having a methylation state associated with a subject who has cervical cancer, a cervical cancer subtype (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., in-situ adenocarcinoma, cervical intraepithelial neoplasia). In some embodiments, the kit comprises a sample collector for obtaining a sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) from a subject. In some embodiments, the kit comprises an oligonucleotide as described herein.

DEFINITIONS

To facilitate an understanding of the present technology, a number of terms and phrases are defined below. Additional definitions are set forth throughout the detailed description.

Throughout the specification and claims, the following terms take the meanings explicitly associated herein, unless the context clearly dictates otherwise. The phrase "in one embodiment" as used herein does not necessarily refer to the same embodiment, though it may. Furthermore, the phrase "in another embodiment" as used herein does not necessarily refer to a different embodiment, although it may. Thus, as described below, various embodiments of the invention may be readily combined, without departing from the scope or spirit of the invention.

In addition, as used herein, the term "or" is an inclusive "or" operator and is equivalent to the term "and/or" unless the context clearly dictates otherwise. The term "based on" is not exclusive and allows for being based on additional factors not described, unless the context clearly dictates otherwise. In addition, throughout the specification, the meaning of "a", "an", and "the" include plural references. The meaning of "in" includes "in" and "on."

The transitional phrase "consisting essentially of" as used in claims in the present application limits the scope of a claim to the specified materials or steps "and those that do

not materially affect the basic and novel characteristic(s)" of the claimed invention, as discussed in *In re Herz, 537 F.2d 549, 551-52, 190 USPQ 461, 463 (CCPA 1976)*. For example, a composition "consisting essentially of" recited elements may contain an unrecited contaminant at a level such that, though present, the contaminant does not alter the function of the recited composition as compared to a pure composition, *i.e.*, a composition "consisting of" the recited components.

The term "one or more", as used herein, refers to a number higher than one. For example, the term "one or more" encompasses any of the following: two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, twelve or more, thirteen or more, fourteen or more, fifteen or more, twenty or more, fifty or more, 100 or more, or an even greater number.

The term "one or more but less than a higher number", "two or more but less than a higher number", "three or more but less than a higher number", "four or more but less than a higher number", "five or more but less than a higher number", "six or more but less than a higher number", "seven or more but less than a higher number", "eight or more but less than a higher number", "nine or more but less than a higher number", "ten or more but less than a higher number", "eleven or more but less than a higher number", "twelve or more but less than a higher number", "thirteen or more but less than a higher number", "fourteen or more but less than a higher number", or "fifteen or more but less than a higher number" is not limited to a higher number. For example, the higher number can be 10,000, 1,000, 100, 50, etc. For example, the higher number can be approximately 64 (e.g., 64, 63, 62, 61, 60, 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 32, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3 or 2). For example, the higher number can be approximately 423.

The term "one or more methylated markers" or "one or more DMRs" or "one or more genes" or "one or more markers" or "a plurality of methylated markers" or "a plurality of markers" or "a plurality of genes" or "a plurality of DMRs" is similarly not limited to a particular numerical combination. Indeed, any numerical combination of methylated markers is contemplated (e.g., 1-2 methylated markers, 1-3, 1-4, 1-5. 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-26, 1-27, 1-28, 1-29, 1-30, 1-31, 1-32, 1-33, 1-34, 1-35, 1-36, 1-37, 1-38, 1-39, 1-40, 1-41, 1-42, 1-43, 1-44, 1-45, 1-46, 1-47, 1-48, 1-49, 1-50, 1-51, 1-52, 1-53, 1-54, 1-55, 1-56, 1-57, 1-58, 1-59, 1-60, 1-61, 1-62, 1-63, 1-64) (e.g., 2-3, 2-4, 2-5, 2-6, 2-7, 2-8, 2-9, 2-10, 2-11, 2-12, 2-13, 2-14,

2-15, 2-16, 2-17, 2-18, 2-19, 2-20, 2-21, 2-22, 2-23, 2-24, 2-25, 2-26, 2-27, 2-28, 2-29, 2-30, 2-31, 2-32, 2-33, 2-34, 2-35, 2-36, 2-37, 2-38, 2-39, 2-40, 2-41, 2-42, 2-43, 2-44, 2-45, 2-46, 2-47, 2-48, 2-49, 2-50, 2-51, 2-52, 2-53, 2-54, 2-55, 2-56, 2-57, 2-58, 2-59, 2-60, 2-61, 2-62, 2-63, 2-64) (e.g., 3-4, 3-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-11, 3-12, 3-13, 3-14, 3-15, 3-16, 3-17, 3-18, 3-19, 3-20, 3-21, 3-22, 3-23, 3-24, 3-25, 3-26, 3-27, 3-28, 3-29, 3-30, 3-31, 3-32, 3-33, 3-34, 3-35, 3-36, 3-37, 3-38, 3-39, 3-40, 3-41, 3-42, 3-43, 3-44, 3-45, 3-46, 3-47, 3-48, 3-49, 3-50, 3-51, 3-52, 3-53, 3-54, 3-55, 3-56, 3-57, 3-58, 3-59, 3-60, 3-61, 3-62, 3-63, 3-64) (e.g., 4-5, 4-6, 4-7, 4-8, 4-9, 4-10, 4-11, 4-12, 4-13, 4-14, 4-15, 4-16, 4-17, 4-18, 4-19, 4-20, 4-21, 4-22, 4-23, 4-24, 4-25, 4-26, 4-27, 4-28, 4-29, 4-30, 4-31, 4-32, 4-33, 4-34, 4-35, 4-36, 4-37, 4-38, 4-39, 4-40, 4-41, 4-42, 4-43, 4-44, 4-45, 4-46, 4-47, 4-48, 4-49, 4-50, 4-51, 4-52, 4-53, 4-54, 4-55, 4-56, 4-57, 4-58, 4-59, 4-60, 4-61, 4-62, 4-63, 4-64) (e.g., 5-6, 5-7, 5-8, 5-9, 5-10, 5-11, 5-12, 5-13, 5-14, 5-15, 5-16, 5-17, 5-18, 5-19, 5-20, 5-21, 5-22, 5-23, 5-24, 5-25, 5-26, 5-27, 5-28, 5-29, 5-30, 5-31, 5-32, 5-33, 5-34, 5-35, 5-36, 5-37, 5-38, 5-39, 5-40, 5-41, 5-42, 5-43, 5-44, 5-45, 5-46, 5-47, 5-48, 5-49, 5-50, 5-51, 5-52, 5-53, 5-54, 5-55, 5-56, 5-57, 5-58, 5-59, 5-60, 5-61, 5-62, 5-63, 5-64) (e.g., 6-7, 6-8, 6-9, 6-10, 6-11, 6-12, 6-13, 6-14, 6-15, 6-16, 6-17, 6-18, 6-19, 6-20, 6-21, 6-22, 6-23, 6-24, 6-25, 6-26, 6-27, 6-28, 6-29, 6-30, 6-31, 6-32, 6-33, 6-34, 6-35, 6-36, 6-37, 6-38, 6-39, 6-40, 6-41, 6-42, 6-43, 6-44, 6-45, 6-46, 6-47, 6-48, 6-49, 6-50, 6-51, 6-52, 6-53, 6-54, 6-55, 6-56, 6-57, 6-58, 6-59, 6-60, 6-61, 6-62, 6-63, 6-64) (e.g., 7-8, 7-9, 7-10, 7-11, 7-12, 7-13, 7-14, 7-15, 7-16, 7-17, 7-18, 7-19, 7-20, 7-21, 7-22, 7-23, 7-24, 7-25, 7-26, 7-27, 7-28, 7-29, 7-30, 7-31, 7-32, 7-33, 7-34, 7-35, 7-36, 7-37, 7-38, 7-39, 7-40, 7-41, 7-42, 7-43, 7-44, 7-45, 7-46, 7-47, 7-48, 7-49, 7-50, 7-51, 7-52, 7-53, 7-54, 7-55, 7-56, 7-57, 7-58, 7-59, 7-60, 7-61, 7-62, 7-63, 7-64) (e.g., 8-9, 8-10, 8-11, 8-12, 8-13, 8-14, 8-15, 8-16, 8-17, 8-18, 8-19, 8-20, 8-21, 8-22, 8-23, 8-24, 8-25, 8-26, 8-27, 8-28, 8-29, 8-30, 8-31, 8-32, 8-33, 8-34, 8-35, 8-36, 8-37, 8-38, 8-39, 8-40, 8-41, 8-42, 8-43, 8-44, 8-45, 8-46, 8-47, 8-48, 8-49, 8-50, 8-51, 8-52, 8-53, 8-54, 8-55, 8-56, 8-57, 8-58, 8-59, 8-60, 8-61, 8-62, 8-63, 8-64) (e.g., 9-10, 9-11, 9-12, 9-13, 9-14, 9-15, 9-16, 9-17, 9-18, 9-19, 9-20, 9-21, 9-22, 9-23, 9-24, 9-25, 9-26, 9-27, 9-28, 9-29, 9-30, 9-31, 9-32, 9-33, 9-34, 9-35, 9-36, 9-37, 9-38, 9-39, 9-40, 9-41, 9-42, 9-43, 9-44, 9-45, 9-46, 9-47, 9-48, 9-49, 9-50, 9-51, 9-52, 9-53, 9-54, 9-55, 9-56, 9-57, 9-58, 9-59, 9-60, 9-61, 9-62, 9-63, 9-64) (e.g., 10-11, 10-12, 10-13, 10-14, 10-15, 10-16, 10-17, 10-18, 10-19, 10-20, 10-21, 10-22, 10-23, 10-24, 10-25, 10-26, 10-27, 10-28, 10-29, 10-30, 10-31, 10-32, 10-33, 10-34, 10-35, 10-36, 10-37, 10-38, 10-39, 10-40, 10-41, 10-42, 10-43, 10-44, 10-45, 10-46, 10-47, 10-48, 10-49, 10-50, 10-51, 10-52, 10-53, 10-54, 10-55, 10-56, 10-57, 10-58, 10-59, 10-60, 10-61, 10-62, 10-63, 10-64) (e.g.,

11-12, 11-13, 11-14, 11-15, 11-16, 11-17, 11-18, 11-19, 11-20, 11-21, 11-22, 11-23, 11-24, 11-25, 11-26, 11-27, 11-28, 11-29, 11-30, 11-31, 11-32, 11-33, 11-34, 11-35, 11-36, 11-37, 11-38, 11-39, 11-40, 11-41, 11-42, 11-43, 11-44, 11-45, 11-46, 11-47, 11-48, 11-49, 11-50, 11-51, 11-52, 11-53, 11-54, 11-55, 11-56, 11-57, 11-58, 11-59, 11-60, 11-61, 11-62, 11-63, 11-64) (e.g., 12-13, 12-14, 12-15, 12-16, 12-17, 12-18, 12-19, 12-20, 12-21, 12-22, 12-23, 12-24, 12-25, 12-26, 12-27, 12-28, 12-29, 12-30, 12-31, 12-32, 12-33, 12-34, 12-35, 12-36, 12-37, 12-38, 12-39, 12-40, 12-41, 12-42, 12-43, 12-44, 12-45, 12-46, 12-47, 12-48, 12-49, 12-50, 12-51, 12-52, 12-53, 12-54, 12-55, 12-56, 12-57, 12-58, 12-59, 12-60, 12-61, 12-62, 12-63, 12-64) (e.g., 13-14, 13-15, 13-16, 13-17, 13-18, 13-19, 13-20, 13-21, 13-22, 13-23, 13-24, 13-25, 13-26, 13-27, 13-28, 13-29, 13-30, 13-31, 13-32, 13-33, 13-34, 13-35, 13-36, 13-37, 13-38, 13-39, 13-40, 13-41, 13-42, 13-43, 13-44, 13-45, 13-46, 13-47, 13-48, 13-49, 13-50, 13-51, 13-52, 13-53, 13-54, 13-55, 13-56, 13-57, 13-58, 13-59, 13-60, 13-61, 13-62, 13-63, 13-64) (e.g., 14-15, 14-16, 14-17, 14-18, 14-19, 14-20, 14-21, 14-22, 14-23, 14-24, 14-25, 14-26, 14-27, 14-28, 14-29, 14-30, 14-31, 14-32, 14-33, 14-34, 14-35, 14-36, 14-37, 14-38, 14-39, 14-40, 14-41, 14-42, 14-43, 14-44, 14-45, 14-46, 14-47, 14-48, 14-49, 14-50, 14-51, 14-52, 14-53, 14-54, 14-55, 14-56, 14-57, 14-58, 14-59, 14-60, 14-61, 14-62, 14-63, 14-64) (e.g., 15-16, 15-17, 15-18, 15-19, 15-20, 15-21, 15-22, 15-23, 15-24, 15-25, 15-26, 15-27, 15-28, 15-29, 15-30, 15-31, 15-32, 15-33, 15-34, 15-35, 15-36, 15-37, 15-38, 15-39, 15-40, 15-41, 15-42, 15-43, 15-44, 15-45, 15-46, 15-47, 15-48, 15-49, 15-50, 15-51, 15-52, 15-53, 15-54, 15-55, 15-56, 15-57, 15-58, 15-59, 15-60, 15-61, 15-62, 15-63, 15-64) (e.g., 16-17, 16-18, 16-19, 16-20, 16-21, 16-22, 16-23, 16-24, 16-25, 16-26, 16-27, 16-28, 16-29, 16-30, 16-31, 16-32, 16-33, 16-34, 16-35, 16-36, 16-37, 16-38, 16-39, 16-40, 16-41, 16-42, 16-43, 16-44, 16-45, 16-46, 16-47, 16-48, 16-49, 16-50, 16-51, 16-52, 16-53, 16-54, 16-55, 16-56, 16-57, 16-58, 16-59, 16-60, 16-61, 16-62, 16-63, 16-64) (e.g., 17-18, 17-19, 17-20, 17-21, 17-22, 17-23, 17-24, 17-25, 17-26, 17-27, 17-28, 17-29, 17-30, 17-31, 17-32, 17-33, 17-34, 17-35, 17-36, 17-37, 17-38, 17-39, 17-40, 17-41, 17-42, 17-43, 17-44, 17-45, 17-46, 17-47, 17-48, 17-49, 17-50, 17-51, 17-52, 17-53, 17-54, 17-55, 17-56, 17-57, 17-58, 17-59, 17-60, 17-61, 17-62, 17-63, 17-64) (e.g., 18-19, 18-20, 18-21, 18-22, 18-23, 18-24, 18-25, 18-26, 18-27, 18-28, 18-29, 18-30, 18-31, 18-32, 18-33, 18-34, 18-35, 18-36, 18-37, 18-38, 18-39, 18-40, 18-41, 18-42, 18-43, 18-44, 18-45, 18-46, 18-47, 18-48, 18-49, 18-50, 18-51, 18-52, 18-53, 18-54, 18-55, 18-56, 18-57, 18-58, 18-59, 18-60, 18-61, 18-62, 18-63, 18-64) (e.g., 19-20, 19-21, 19-22, 19-23, 19-24, 19-25, 19-26, 19-27, 19-28, 19-29, 19-30, 19-31, 19-32, 19-33, 19-34, 19-35, 19-36, 19-37, 19-38, 19-39, 19-40, 19-41, 19-42, 19-43, 19-44, 19-

45, 19-46, 19-47, 19-48, 19-49, 19-50, 19-51, 19-52, 19-53, 19-54, 19-55, 19-56, 19-57, 19-58, 19-59, 19-60, 19-61, 19-62, 19-63, 19-64) (e.g., 20-21, 20-22, 20-23, 20-24, 20-25, 20-26, 20-27, 20-28, 20-29, 20-30, 20-31, 20-32, 20-33, 20-34, 20-35, 20-36, 20-37, 20-38, 20-39, 20-40, 20-41, 20-42, 20-43, 20-44, 20-45, 20-46, 20-47, 20-48, 20-49, 20-50, 20-51, 20-52, 20-53, 20-54, 20-55, 20-56, 20-57, 20-58, 20-59, 20-60, 20-61, 20-62, 20-63, 20-64) (e.g., 21-22, 21-23, 21-24, 21-25, 21-26, 21-27, 21-28, 21-29, 21-30, 21-31, 21-32, 21-33, 21-34, 21-35, 21-36, 21-37, 21-38, 21-39, 21-40, 21-41, 21-42, 21-43, 21-44, 21-45, 21-46, 21-47, 21-48, 21-49, 21-50, 21-51, 21-52, 21-53, 21-54, 21-55, 21-56, 21-57, 21-58, 21-59, 21-60, 21-61, 21-62, 21-63, 21-64) (e.g., 22-23, 22-24, 22-25, 22-26, 22-27, 22-28, 22-29, 22-30, 22-31, 22-32, 22-33, 22-34, 22-35, 22-36, 22-37, 22-38, 22-39, 22-40, 22-41, 22-42, 22-43, 22-44, 22-45, 22-46, 22-47, 22-48, 22-49, 22-50, 22-51, 22-52, 22-53, 22-54, 22-55, 22-56, 22-57, 22-58, 22-59, 22-60, 22-61, 22-62, 22-63, 22-64) (e.g., 23-24, 23-25, 23-26, 23-27, 23-28, 23-29, 23-30, 23-31, 23-32, 23-33, 23-34, 23-35, 23-36, 23-37, 23-38, 23-39, 23-40, 23-41, 23-42, 23-43, 23-44, 23-45, 23-46, 23-47, 23-48, 23-49, 23-50, 23-51, 23-52, 23-53, 23-54, 23-55, 23-56, 23-57, 23-58, 23-59, 23-60, 23-61, 23-62, 23-63, 23-64) (e.g., 24-25, 24-26, 24-27, 24-28, 24-29, 24-30, 24-31, 24-32, 24-33, 24-34, 24-35, 24-36, 24-37, 24-38, 24-39, 24-40, 24-41, 24-42, 24-43, 24-44, 24-45, 24-46, 24-47, 24-48, 24-49, 24-50, 24-51, 24-52, 24-53, 24-54, 24-55, 24-56, 24-57, 24-58, 24-59, 24-60, 24-61, 24-62, 24-63, 24-64) (e.g., 25-26, 25-27, 25-28, 25-29, 25-30, 25-31, 25-32, 25-33, 25-34, 25-35, 25-36, 25-37, 25-38, 25-39, 25-40, 25-41, 25-42, 25-43, 25-44, 25-45, 25-46, 25-47, 25-48, 25-49, 25-50, 25-51, 25-52, 25-53, 25-54, 25-55, 25-56, 25-57, 25-58, 25-59, 25-60, 25-61, 25-62, 25-63, 25-64) (e.g., 26-27, 26-28, 26-29, 26-30, 26-31, 26-32, 26-33, 26-34, 26-35, 26-36, 26-37, 26-38, 26-39, 26-40, 26-41, 26-42, 26-43, 26-44, 26-45, 26-46, 26-47, 26-48, 26-49, 26-50, 26-51, 26-52, 26-53, 26-54, 26-55, 26-56, 26-57, 26-58, 26-59, 26-60, 26-61, 26-62, 26-63, 26-64) (e.g., 27-28, 27-29, 27-30, 27-31, 27-32, 27-33, 27-34, 27-35, 27-36, 27-37, 27-38, 27-39, 27-40, 27-41, 27-42, 27-43, 27-44, 27-45, 27-46, 27-47, 27-48, 27-49, 27-50, 27-51, 27-52, 27-53, 27-54, 27-55, 27-56, 27-57, 27-58, 27-59, 27-60, 27-61, 27-62, 27-63, 27-64) (e.g., 28-29, 28-30, 28-31, 28-32, 28-33, 28-34, 28-35, 28-36, 28-37, 28-38, 28-39, 28-40, 28-41, 28-42, 28-43, 28-44, 28-45, 28-46, 28-47, 28-48, 28-49, 28-50, 28-51, 28-52, 28-53, 28-54, 28-55, 28-56, 28-57, 28-58, 28-59, 28-60, 28-61, 28-62, 28-63, 28-64) (e.g., 29-30, 29-31, 29-32, 29-33, 29-34, 29-35, 29-36, 29-37, 29-38, 29-39, 29-40, 29-41, 29-42, 29-43, 29-44, 29-45, 29-46, 29-47, 29-48, 29-49, 29-50, 29-51, 29-52, 29-53, 29-54, 29-55, 29-56, 29-57, 29-58, 29-59, 29-60, 29-61, 29-62, 29-63, 29-64) (e.g., 30-31, 30-32, 30-33, 30-34, 30-35, 30-36,

30-37, 30-38, 30-39, 30-40, 30-41, 30-42, 30-43, 30-44, 30-45, 30-46, 30-47, 30-48, 30-49, 30-50, 30-51, 30-52, 30-53, 30-54, 30-55, 30-56, 30-57, 30-58, 30-59, 30-60, 30-61, 30-62, 30-63, 30-64) (e.g., 31-32, 31-33, 31-34, 31-35, 31-36, 31-37, 31-38, 31-39, 31-40, 31-41, 31-42, 31-43, 31-44, 31-45, 31-46, 31-47, 31-48, 31-49, 31-50, 31-51, 31-52, 31-53, 31-54, 31-55, 31-56, 31-57, 31-58, 31-59, 31-60, 31-61, 31-62, 31-63, 31-64) (e.g., 32-33, 32-34, 32-35, 32-36, 32-37, 32-38, 32-39, 32-40, 32-41, 32-42, 32-43, 32-44, 32-45, 32-46, 32-47, 32-48, 32-49, 32-50, 32-51, 32-52, 32-53, 32-54, 32-55, 32-56, 32-57, 32-58, 32-59, 32-60, 32-61, 32-62, 32-63, 32-64) (e.g., 33-34, 33-35, 33-36, 33-37, 33-38, 33-39, 33-40, 33-41, 33-42, 33-43, 33-44, 33-45, 33-46, 33-47, 33-48, 33-49, 33-50, 33-51, 33-52, 33-53, 33-54, 33-55, 33-56, 33-57, 33-58, 33-59, 33-60, 33-61, 33-62, 33-63, 33-64) (e.g., 34-35, 34-36, 34-37, 34-38, 34-39, 34-40, 34-41, 34-42, 34-43, 34-44, 34-45, 34-46, 34-47, 34-48, 34-49, 34-50, 34-51, 34-52, 34-53, 34-54, 34-55, 34-56, 34-57, 34-58, 34-59, 34-60, 34-61, 34-62, 34-63, 34-64) (e.g., 35-36, 35-37, 35-38, 35-39, 35-40, 35-41, 35-42, 35-43, 35-44, 35-45, 35-46, 35-47, 35-48, 35-49, 35-50, 35-51, 35-52, 35-53, 35-54, 35-55, 35-56, 35-57, 35-58, 35-59, 35-60, 35-61, 35-62, 35-63, 35-64) (e.g., 36-37, 36-38, 36-39, 36-40, 36-41, 36-42, 36-43, 36-44, 36-45, 36-46, 36-47, 36-48, 36-49, 36-50, 36-51, 36-52, 36-53, 36-54, 36-55, 36-56, 36-57, 36-58, 36-59, 36-60, 36-61, 36-62, 36-63, 36-64) (e.g., 37-38, 37-39, 37-40, 37-41, 37-42, 37-43, 37-44, 37-45, 37-46, 37-47, 37-48, 37-49, 37-50, 37-51, 37-52, 37-53, 37-54, 37-55, 37-56, 37-57, 37-58, 37-59, 37-60, 37-61, 37-62, 37-63, 37-64) (e.g., 38-39, 38-40, 38-41, 38-42, 38-43, 38-44, 38-45, 38-46, 38-47, 38-48, 38-49, 38-50, 38-51, 38-52, 38-53, 38-54, 38-55, 38-56, 38-57, 38-58, 38-59, 38-60, 38-61, 38-62, 38-63, 38-64), (e.g., 39-40, 39-41, 39-42, 39-43, 39-44, 39-45, 39-46, 39-47, 39-48, 39-49, 39-50, 39-51, 39-52, 39-53, 39-54, 39-55, 39-56, 39-57, 39-58, 39-59, 39-60, 39-61, 39-62, 39-63, 39-64), (e.g., 40-41, 40-42, 40-43, 40-44, 40-45, 40-46, 40-47, 40-48, 40-49, 40-50, 40-51, 40-52, 40-53, 40-54, 40-55, 40-56, 40-57, 40-58, 40-59, 40-60, 40-61, 40-62, 40-63, 40-64), (e.g., 41-42, 41-43, 41-44, 41-45, 41-46, 41-47, 41-48, 41-49, 41-50, 41-51, 41-52, 41-53, 41-54, 41-55, 41-56, 41-57, 41-58, 41-59, 41-60, 41-61, 41-62, 41-63, 41-64), (e.g., 42-43, 42-44, 42-45, 42-46, 42-47, 42-48, 42-49, 42-50, 42-51, 42-52, 42-53, 42-54, 42-55, 42-56, 42-57, 42-58, 42-59, 42-60, 42-61, 42-62, 42-63, 42-64) (e.g., 43-44, 43-45, 43-46, 43-47, 43-48, 43-49, 43-50, 43-51, 43-52, 43-53, 43-54, 43-55, 43-56, 43-57, 43-58, 43-59, 43-60, 43-61, 43-62, 43-63, 43-64) (e.g., 44-45, 44-46, 44-47, 44-48, 44-49, 44-50, 44-51, 44-52, 44-53, 44-54, 44-55, 44-56, 44-57, 44-58, 44-59, 44-60, 44-61, 44-62, 44-63, 44-64) (e.g., 45-46, 45-47, 45-48, 45-49, 45-50, 45-51, 45-52, 45-53, 45-54, 45-55, 45-56, 45-57, 45-58, 45-59, 45-60,

45-61, 45-62, 45-63, 45-64) (e.g., 46-47, 46-48, 46-49, 46-50, 46-51, 46-52, 46-53, 46-54, 46-55, 46-56, 46-57, 46-58, 46-59, 46-60, 46-61, 46-62, 46-63, 46-64) (e.g., 47-48, 47-49, 47-50, 47-51, 47-52, 47-53, 47-54, 47-55, 47-56, 47-57, 47-58, 47-59, 47-60, 47-61, 47-62, 47-63, 47-64) (e.g., 48-49, 48-50, 48-51, 48-52, 48-53, 48-54, 48-55, 48-56, 48-57, 48-58, 48-59, 48-60, 48-61, 48-62, 48-63, 48-64) (e.g., 49-50, 49-51, 49-52, 49-53, 49-54, 49-55, 49-56, 49-57, 49-58, 49-59, 49-60, 49-61, 49-62, 49-63, 49-64) (e.g., 50-51, 50-52, 50-53, 50-54, 50-55, 50-56, 50-57, 50-58, 50-59, 50-60, 50-61, 50-62, 50-63, 50-64) (e.g., 51-52, 51-53, 51-54, 51-55, 51-56, 51-57, 51-58, 51-59, 51-60, 51-61, 51-62, 51-63, 51-64) (e.g., 52-53, 52-54, 52-55, 52-56, 52-57, 52-58, 52-59, 52-60, 52-61, 52-62, 52-63, 52-64) (e.g., 53-54, 53-55, 53-56, 53-57, 53-58, 53-59, 53-60, 53-61, 53-62, 53-63 53-64) (e.g., 54-55, 54-56, 54-57, 54-58, 54-59, 54-60, 54-61, 54-62, 54-63, 54-64) (e.g., 55-56, 55-57, 55-58, 55-59, 55-60, 55-61, 55-62, 55-63, 55-64) (e.g., 56-57, 56-58, 56-59, 56-60, 56-61, 56-62, 56-63, 56-64) (e.g., 57-58, 57-59, 57-60, 57-61, 57-62, 57-63, 57-64) (e.g., 58-59, 58-60, 58-61, 58-62, 58-63, 58-64) (e.g., 59-60, 59-61, 59-62, 59-63, 59-64) (e.g., 60-61, 60-62, 60-63, 60-64) (e.g., 61-62, 61-63, 61-64) (e.g., 62-63, 62-64) (e.g., 63-64) (e.g., 1-423, 10-423, 20-423, 41-423, 50-423, 62-423, 64-423, 100-423, 150-423, 200-423, 300-423, 320-423, 361-423, 400-423, 410-423, 415-423, 420-423, 421-423, 422-423) (e.g., 423 or fewer, 400 or fewer, 350 or fewer, 361 or fewer, 320 or fewer, 300 or fewer, 250 or fewer, 200 or fewer, 150 or fewer, 100 or fewer, 75 or fewer, 70 or fewer, 65 or fewer, 64 or fewer, 62 or fewer; 61 or fewer; 60 or fewer; 59 or fewer; 58 or fewer; 57 or fewer; 56 or fewer; 55 or fewer; 54 or fewer; 53 or fewer; 52 or fewer; 51 or fewer; 50 or fewer; 49 or fewer; 48 or fewer; 47 or fewer; 46 or fewer; 45 or fewer; 44 or fewer; 43 or fewer; 42 or fewer; 41 or fewer; 40 or fewer; 39 or fewer; 38 or fewer; 37 or fewer; 36 or fewer; 35 or fewer; 34 or fewer; 33 or fewer; 32 or fewer; 31 or fewer; 30 or fewer; 29 or fewer; 28 or fewer; 27 or fewer; 26 or fewer; 25 or fewer; 24 or fewer; 23 or fewer; 22 or fewer; 21 or fewer; 20 or fewer; 19 or fewer; 18 or fewer; 17 or fewer; 16 or fewer; 15 or fewer; 14 or fewer; 13 or fewer; 12 or fewer; 11 or fewer; 10 or fewer; 9 or fewer; 8 or fewer; 7 or fewer; 6 or fewer; 5 or fewer; 4 or fewer; 3 or fewer; 2 or 1).

As used herein, a "nucleic acid" or "nucleic acid molecule" generally refers to any ribonucleic acid or deoxyribonucleic acid, which may be unmodified or modified DNA or RNA. "Nucleic acids" include, without limitation, single- and double-stranded nucleic acids. As used herein, the term "nucleic acid" also includes DNA as described above that contains one or more modified bases. Thus, DNA with a backbone modified for stability or for other

reasons is a "nucleic acid". The term "nucleic acid" as it is used herein embraces such chemically, enzymatically, or metabolically modified forms of nucleic acids, as well as the chemical forms of DNA characteristic of viruses and cells, including for example, simple and complex cells.

The terms "oligonucleotide" or "polynucleotide" or "nucleotide" or "nucleic acid" refer to a molecule having two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof. Typical deoxyribonucleotides for DNA are thymine, adenine, cytosine, and guanine. Typical ribonucleotides for RNA are uracil, adenine, cytosine, and guanine.

As used herein, the terms "locus" or "region" of a nucleic acid refer to a subregion of a nucleic acid, e.g., a gene on a chromosome, a single nucleotide, a CpG island, etc.

The terms "complementary" and "complementarity" refer to nucleotides (e.g., 1 nucleotide) or polynucleotides (e.g., a sequence of nucleotides) related by the base-pairing rules. For example, the sequence 5'-A-G-T-3' is complementary to the sequence 3'-T-C-A-5'. Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands effects the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions and in detection methods that depend upon binding between nucleic acids.

The term "gene" refers to a nucleic acid (e.g., DNA or RNA) sequence that comprises coding sequences necessary for the production of an RNA, or of a polypeptide or its precursor. A functional polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence as long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction, etc.) of the polypeptide are retained. The term "portion" when used in reference to a gene refers to fragments of that gene. The fragments may range in size from a few nucleotides to the entire gene sequence minus one nucleotide. Thus, "a nucleotide comprising at least a portion of a gene" may comprise fragments of the gene or the entire gene.

The term "gene" also encompasses the coding regions of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends, e.g., for a distance of about 1 kb on either end, such that the gene corresponds to the length of the full-length mRNA (e.g., comprising coding, regulatory, structural and other sequences). The sequences that are located 5' of the coding region and that are present on the mRNA are referred to as 5' non-translated or untranslated sequences. The sequences that are located 3' or downstream of the coding region and that are present on the mRNA are referred to as 3' non-translated or 3' untranslated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. In some organisms (e.g., eukaryotes), a genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' ends of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, posttranscriptional cleavage, and polyadenylation.

The term "wild-type" when made in reference to a gene refers to a gene that has the characteristics of a gene isolated from a naturally occurring source. The term "wild-type" when made in reference to a gene product refers to a gene product that has the characteristics of a gene product isolated from a naturally occurring source. The term "wild-type" when made in reference to a protein refers to a protein that has the characteristics of a naturally occurring protein. The term "naturally-occurring" as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by the hand of a person in the laboratory is naturally-occurring. A wild-type gene is often that gene or allele that is most frequently

observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" when made in reference to a gene or to a gene product refers, respectively, to a gene or to a gene product that displays modifications in sequence and/or functional properties (e.g., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term "allele" refers to a variation of a gene; the variations include but are not limited to variants and mutants, polymorphic loci, and single nucleotide polymorphic loci, frameshift, and splice mutations. An allele may occur naturally in a population or it might arise during the lifetime of any particular individual of the population.

Thus, the terms "variant" and "mutant" when used in reference to a nucleotide sequence refer to a nucleic acid sequence that differs by one or more nucleotides from another, usually related, nucleotide acid sequence. A "variation" is a difference between two different nucleotide sequences; typically, one sequence is a reference sequence.

The term "primer" refers to an oligonucleotide, whether occurring naturally as, *e.g.*, a nucleic acid fragment from a restriction digest, or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid template strand is induced, (*e.g.*, in the presence of nucleotides and an inducing agent such as a DNA polymerase, and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer, and the use of the method. In some embodiments, the primer pair is specific for a specific MDM (e.g., MDMs in Tables I, III, and X) and specifically binds at least a portion of a genetic region comprising the MDM (e.g., chromosomal coordinates in Tables I, III and/or X).

The term "probe" refers to an oligonucleotide (*e.g.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly, or by PCR amplification, that is capable of hybridizing to another

oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification, and isolation of particular gene sequences (*e.g.*, a "capture probe"). It is contemplated that any probe used in the present invention may, in some embodiments, be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The term "target," as used herein refers to a nucleic acid sought to be sorted out from other nucleic acids, *e.g.*, by probe binding, amplification, isolation, capture, *etc.* For example, when used in reference to the polymerase chain reaction, "target" refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction, while when used in an assay in which target DNA is not amplified, *e.g.*, in some embodiments of an invasive cleavage assay, a target comprises the site at which a probe and invasive oligonucleotides (*e.g.*, INVADER oligonucleotide) bind to form an invasive cleavage structure, such that the presence of the target nucleic acid can be detected. A "segment" is defined as a region of nucleic acid within the target sequence.

Accordingly, as used herein, "non-target", e.g., as it is used to describe a nucleic acid such as a DNA, refers to nucleic acid that may be present in a reaction, but that is not the subject of detection or characterization by the reaction. In some embodiments, non-target nucleic acid may refer to nucleic acid present in a sample that does not, e.g., contain a target sequence, while in some embodiments, non-target may refer to exogenous nucleic acid, i.e., nucleic acid that does not originate from a sample containing or suspected of containing a target nucleic acid, and that is added to a reaction, e.g., to normalize the activity of an enzyme (e.g., polymerase) to reduce variability in the performance of the enzyme in the reaction.

As used herein, "methylation" refers to cytosine methylation at positions C5 or N4 of cytosine, the N6 position of adenine, or other types of nucleic acid methylation. In vitro amplified DNA is usually unmethylated because typical in vitro DNA amplification methods do not retain the methylation pattern of the amplification template. However, "unmethylated DNA" or "methylated DNA" can also refer to amplified DNA whose original template was unmethylated or methylated, respectively.

As used herein, the term "amplification reagents" refers to those reagents (deoxyribonucleoside triphosphates, buffer, etc.), needed for amplification except for primers,

nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel.

As used herein, the term "control" when used in reference to nucleic acid detection or analysis refers to a nucleic acid having known features (e.g., known sequence, known copy-number per cell), for use in comparison to an experimental target (e.g., a nucleic acid of unknown concentration). A control may be an endogenous, preferably invariant gene against which a test or target nucleic acid in an assay can be normalized. Such normalizing controls for sample-to-sample variations that may occur in, for example, sample processing, assay efficiency, etc., and allows accurate sample-to-sample data comparison. Genes that find use for normalizing nucleic acid detection assays on human samples include, e.g., β-actin, ZDHHC1, and B3GALT6 (see, e.g., U.S. patent application Ser. Nos 14/966,617 and 62/364,082, each incorporated herein by reference). As used herein "ZDHHC1" refers to a gene encoding a protein characterized as a zinc finger, DHHC-type containing 1, located in human DNA on Chr 16 (16q22.1) and belonging to the DHHC palmitoyltransferase family.

Controls may also be external. For example, in quantitative assays such as qPCR, QuARTS, etc., a "calibrator" or "calibration control" is a nucleic acid of known sequence, e.g., having the same sequence as a portion of an experimental target nucleic acid, and a known concentration or series of concentrations (e.g., a serially diluted control target for generation of calibration curved in quantitative PCR). Typically, calibration controls are analyzed using the same reagents and reaction conditions as are used on an experimental DNA. In certain embodiments, the measurement of the calibrators is done at the same time, e.g., in the same thermal cycler, as the experimental assay. In preferred embodiments, multiple calibrators may be included in a single plasmid, such that the different calibrator sequences are easily provided in equimolar amounts. In particularly preferred embodiments, plasmid calibrators are digested, e.g., with one or more restriction enzymes, to release calibrator portion from the plasmid vector. See, e.g., WO 2015/066695, which is included herein by reference.

As used herein a “methylated nucleotide” or a “methylated nucleotide base” refers to the presence of a methyl moiety on a nucleotide base, where the methyl moiety is not present in a recognized typical nucleotide base. For example, cytosine does not contain a methyl moiety on its pyrimidine ring, but 5-methylcytosine contains a methyl moiety at position 5 of its pyrimidine ring. Therefore, cytosine is not a methylated nucleotide and 5-methylcytosine is a methylated nucleotide. In another example, thymine contains a methyl moiety at position

5 of its pyrimidine ring; however, for purposes herein, thymine is not considered a methylated nucleotide when present in DNA since thymine is a typical nucleotide base of DNA.

As used herein, a "methylated nucleic acid molecule" refers to a nucleic acid molecule that contains one or more methylated nucleotides.

As used herein, a "methylation state", "methylation profile", and "methylation status" of a nucleic acid molecule refers to the presence or absence of one or more methylated nucleotide bases in the nucleic acid molecule. For example, a nucleic acid molecule containing a methylated cytosine is considered methylated (e.g., the methylation state of the nucleic acid molecule is methylated). A nucleic acid molecule that does not contain any methylated nucleotides is considered unmethylated.

As used herein, the term "methylation level" as applied to a methylation marker refers to the amount of methylation within a particular methylation marker. Methylation level may also refer to the amount of methylation within a particular methylation marker in comparison with an established norm or control. Methylation level may also refer to whether one or more cytosine residues present in a CpG context have or do not have a methylation group. Methylation level may also refer to the fraction of cells in a sample that do or do not have a methylation group on such cytosines. Methylation level may also alternatively describe whether a single CpG di-nucleotide is methylated.

The methylation state of a particular nucleic acid sequence (e.g., a gene marker or DNA region as described herein) can indicate the methylation state of every base in the sequence or can indicate the methylation state of a subset of the bases (e.g., of one or more cytosines) within the sequence, or can indicate information regarding regional methylation density within the sequence with or without providing precise information of the locations within the sequence the methylation occurs.

The methylation state of a nucleotide locus in a nucleic acid molecule refers to the presence or absence of a methylated nucleotide at a particular locus in the nucleic acid molecule. For example, the methylation state of a cytosine at the 7th nucleotide in a nucleic acid molecule is methylated when the nucleotide present at the 7th nucleotide in the nucleic acid molecule is 5-methylcytosine. Similarly, the methylation state of a cytosine at the 7th nucleotide in a nucleic acid molecule is unmethylated when the nucleotide present at the 7th nucleotide in the nucleic acid molecule is cytosine (and not 5-methylcytosine).

The methylation status can optionally be represented or indicated by a "methylation value" (e.g., representing a methylation frequency, fraction, ratio, percent, etc.). A methylation value can be generated, for example, by quantifying the amount of intact nucleic acid present following restriction digestion with a methylation dependent restriction enzyme or by comparing amplification profiles after bisulfite reaction or by comparing sequences of bisulfite-treated and untreated nucleic acids or by comparing TET-treated and untreated nucleic acids. Accordingly, a value, e.g., a methylation value, represents the methylation status and can thus be used as a quantitative indicator of methylation status across multiple copies of a locus. This is of particular use when it is desirable to compare the methylation status of a sequence in a sample to a threshold or reference value.

As used herein, "methylation frequency" or "methylation percent (%)" refer to the number of instances in which a molecule or locus is methylated relative to the number of instances the molecule or locus is unmethylated.

The term "methylation score" as used herein is a score indicative of detected methylation events in a marker or panel of markers in comparison with median methylation events for the marker or panel of markers from a random population of mammals (e.g., a random population of 10, 20, 30, 40, 50, 100, or 500 mammals) that do not have a specific neoplasm of interest. An elevated methylation score in a marker or panel of markers can be any score provided that the score is greater than a corresponding reference score. For example, an elevated score of methylation in a marker or panel of markers can be 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more fold greater than the reference methylation score.

As such, the methylation state describes the state of methylation of a nucleic acid (e.g., a genomic sequence). In addition, the methylation state refers to the characteristics of a nucleic acid segment at a particular genomic locus relevant to methylation. Such characteristics include, but are not limited to, whether any of the cytosine (C) residues within this DNA sequence are methylated, the location of methylated C residue(s), the frequency or percentage of methylated C throughout any particular region of a nucleic acid, and allelic differences in methylation due to, e.g., difference in the origin of the alleles. The terms "methylation state", "methylation profile", and "methylation status" also refer to the relative concentration, absolute concentration, or pattern of methylated C or unmethylated C throughout any particular region of a nucleic acid in a biological sample. For example, if the cytosine (C) residue(s) within a nucleic acid sequence are methylated it may be referred to as "hypermethylated" or having "increased methylation", whereas if the cytosine (C) residue(s)

within a DNA sequence are not methylated it may be referred to as "hypomethylated" or having "decreased methylation". Likewise, if the cytosine (C) residue(s) within a nucleic acid sequence are methylated as compared to another nucleic acid sequence (e.g., from a different region or from a different individual, etc.) that sequence is considered hypermethylated or having increased methylation compared to the other nucleic acid sequence. Alternatively, if the cytosine (C) residue(s) within a DNA sequence are not methylated as compared to another nucleic acid sequence (e.g., from a different region or from a different individual, etc.) that sequence is considered hypomethylated or having decreased methylation compared to the other nucleic acid sequence. Additionally, the term "methylation pattern" as used herein refers to the collective sites of methylated and unmethylated nucleotides over a region of a nucleic acid. Two nucleic acids may have the same or similar methylation frequency or methylation percent but have different methylation patterns when the number of methylated and unmethylated nucleotides are the same or similar throughout the region but the locations of methylated and unmethylated nucleotides are different. Sequences are said to be "differentially methylated" or as having a "difference in methylation" or having a "different methylation state" when they differ in the extent (e.g., one has increased or decreased methylation relative to the other), frequency, or pattern of methylation. The term "differential methylation" refers to a difference in the level or pattern of nucleic acid methylation in a cancer positive sample as compared with the level or pattern of nucleic acid methylation in a cancer negative sample. It may also refer to the difference in levels or patterns between patients that have recurrence of cancer after surgery versus patients who not have recurrence. Differential methylation and specific levels or patterns of DNA methylation are prognostic and predictive biomarkers, e.g., once the correct cut-off or predictive characteristics have been defined.

Methylation state frequency can be used to describe a population of individuals or a sample from a single individual. For example, a nucleotide locus having a methylation state frequency of 50% is methylated in 50% of instances and unmethylated in 50% of instances. Such a frequency can be used, for example, to describe the degree to which a nucleotide locus or nucleic acid region is methylated in a population of individuals or a collection of nucleic acids. Thus, when methylation in a first population or pool of nucleic acid molecules is different from methylation in a second population or pool of nucleic acid molecules, the methylation state frequency of the first population or pool will be different from the methylation state frequency of the second population or pool. Such a frequency also can be

used, for example, to describe the degree to which a nucleotide locus or nucleic acid region is methylated in a single individual. For example, such a frequency can be used to describe the degree to which a group of cells from a tissue sample are methylated or unmethylated at a nucleotide locus or nucleic acid region.

Typically, methylation of human DNA occurs on a dinucleotide sequence including an adjacent guanine and cytosine where the cytosine is located 5' of the guanine (also termed CpG dinucleotide sequences). Most cytosines within the CpG dinucleotides are methylated in the human genome, however some remain unmethylated in specific CpG dinucleotide rich genomic regions, known as CpG islands (see, e.g, Antequera et al. (1990) *Cell* **62**: 503–514).

As used herein, a "CpG island" or "cytosine-phosphate-guanine island") refers to a G:C-rich region of genomic DNA containing an increased number of CpG dinucleotides relative to total genomic DNA. A CpG island can be at least 100, 200, or more base pairs in length, where the G:C content of the region is at least 50% and the ratio of observed CpG frequency over expected frequency is 0.6; in some instances, a CpG island can be at least 500 base pairs in length, where the G:C content of the region is at least 55%) and the ratio of observed CpG frequency over expected frequency is 0.65. The observed CpG frequency over expected frequency can be calculated according to the method provided in Gardiner-Garden et al (1987) J. Mol. Biol. 196: 261–281. For example, the observed CpG frequency over expected frequency can be calculated according to the formula $R = (A \times B) / (C \times D)$, where R is the ratio of observed CpG frequency over expected frequency, A is the number of CpG dinucleotides in an analyzed sequence, B is the total number of nucleotides in the analyzed sequence, C is the total number of C nucleotides in the analyzed sequence, and D is the total number of G nucleotides in the analyzed sequence. Methylation state is typically determined in CpG islands, e.g., at promoter regions. It will be appreciated though that other sequences in the human genome are prone to DNA methylation such as CpA and CpT (see Ramsahoye (2000) Proc. Natl. Acad. Sci. USA 97: 5237–5242; Salmon and Kaye (1970) Biochim. Biophys. Acta. 204: 340–351; Grafstrom (1985) Nucleic Acids Res. 13: 2827–2842; Nyce (1986) Nucleic Acids Res. 14: 4353–4367; Woodcock (1987) Biochem. Biophys. Res. Commun. 145: 888-894).

As used herein, a "methylation-specific reagent" refers to a reagent that modifies a nucleotide of the nucleic acid molecule as a function of the methylation state of the nucleic acid molecule, or a methylation-specific reagent, refers to a compound or composition or other agent that can change the nucleotide sequence of a nucleic acid molecule in a manner

that reflects the methylation state of the nucleic acid molecule. Methods of treating a nucleic acid molecule with such a reagent can include contacting the nucleic acid molecule with the reagent, coupled with additional steps, if desired, to accomplish the desired change of nucleotide sequence. Such methods can be applied in a manner in which unmethylated nucleotides (*e.g.*, each unmethylated cytosine) is modified to a different nucleotide. For example, in some embodiments, such a reagent can deaminate unmethylated cytosine nucleotides to produce deoxy uracil residues. Examples of such reagents include, but are not limited to, a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, a bisulfite reagent, a TET enzyme, and a borane reducing agent.

A change in the nucleic acid nucleotide sequence by a methylation –specific reagent can also result in a nucleic acid molecule in which each methylated nucleotide is modified to a different nucleotide.

The term "methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of a nucleic acid.

The term "MS AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, as described by Gonzalgo et al. (1997) *Cancer Research* **57**: 594–599.

The term "MethyLight™" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al. (1999) *Cancer Res.* **59**: 2302–2306.

The term "HeavyMethyl™" refers to an assay wherein methylation specific blocking probes (also referred to herein as blockers) covering CpG positions between, or covered by, the amplification primers enable methylation-specific selective amplification of a nucleic acid sample.

The term "HeavyMethyl™ MethyLight™" assay refers to a HeavyMethyl™ MethyLight™ assay, which is a variation of the MethyLight™ assay, wherein the MethyLight™ assay is combined with methylation specific blocking probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones (1997) *Nucleic Acids Res.* 25: 2529–2531.

The term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. (1996) *Proc. Natl. Acad. Sci. USA* **93**: 9821–9826, and by U.S. Pat. No. 5,786,146.

The term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird (1997) *Nucleic Acids Res.* **25**: 2532–2534.

The term "MCA" (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al. (1999) *Cancer Res.* **59**: 2307–12, and in WO 00/26401A1.

As used herein, a "selected nucleotide" refers to one nucleotide of the four typically occurring nucleotides in a nucleic acid molecule (C, G, T, and A for DNA and C, G, U, and A for RNA), and can include methylated derivatives of the typically occurring nucleotides (e.g., when C is the selected nucleotide, both methylated and unmethylated C are included within the meaning of a selected nucleotide), whereas a methylated selected nucleotide refers specifically to a methylated typically occurring nucleotide and an unmethylated selected nucleotides refers specifically to an unmethylated typically occurring nucleotide.

The term "methylation-specific restriction enzyme" refers to a restriction enzyme that selectively digests a nucleic acid dependent on the methylation state of its recognition site. In the case of a restriction enzyme that specifically cuts if the recognition site is not methylated or is hemi-methylated (a methylation-sensitive enzyme), the cut will not take place (or will take place with a significantly reduced efficiency) if the recognition site is methylated on one or both strands. In the case of a restriction enzyme that specifically cuts only if the recognition site is methylated (a methylation-dependent enzyme), the cut will not take place (or will take place with a significantly reduced efficiency) if the recognition site is not methylated. Preferred are methylation-specific restriction enzymes, the recognition sequence of which contains a CG dinucleotide (for instance a recognition sequence such as CGCG or CCCGGG). Further preferred for some embodiments are restriction enzymes that do not cut if the cytosine in this dinucleotide is methylated at the carbon atom C5.

As used herein, the "sensitivity" of a given marker (or set of markers used together) refers to the percentage of samples that report a DNA methylation value above a threshold value that distinguishes between neoplastic and non-neoplastic samples. In some embodiments, a positive is defined as a histology-confirmed neoplasia that reports a DNA methylation value above a threshold value (*e.g.*, the range associated with disease), and a false negative is defined as a histology-confirmed neoplasia that reports a DNA methylation

value below the threshold value (*e.g.*, the range associated with no disease). The value of sensitivity, therefore, reflects the probability that a DNA methylation measurement for a given marker obtained from a known diseased sample will be in the range of disease-associated measurements. As defined here, the clinical relevance of the calculated sensitivity value represents an estimation of the probability that a given marker would detect the presence of a clinical condition when applied to a subject with that condition.

As used herein, the "specificity" of a given marker (or set of markers used together) refers to the percentage of non-neoplastic samples that report a DNA methylation value below a threshold value that distinguishes between neoplastic and non-neoplastic samples. In some embodiments, a negative is defined as a histology-confirmed non-neoplastic sample that reports a DNA methylation value below the threshold value (*e.g.*, the range associated with no disease) and a false positive is defined as a histology-confirmed non-neoplastic sample that reports a DNA methylation value above the threshold value (*e.g.*, the range associated with disease). The value of specificity, therefore, reflects the probability that a DNA methylation measurement for a given marker obtained from a known non-neoplastic sample will be in the range of non-disease associated measurements. As defined here, the clinical relevance of the calculated specificity value represents an estimation of the probability that a given marker would detect the absence of a clinical condition when applied to a patient without that condition.

The term "AUC" as used herein is an abbreviation for the "area under a curve". In particular it refers to the area under a Receiver Operating Characteristic (ROC) curve. The ROC curve is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test. It shows the trade-off between sensitivity and specificity depending on the selected cut point (any increase in sensitivity will be accompanied by a decrease in specificity). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test (the larger the area the better; the optimum is 1; a random test would have a ROC curve lying on the diagonal with an area of 0.5; for reference: J. P. Egan. (1975) *Signal Detection Theory and ROC Analysis*, Academic Press, New York).

The term "neoplasm" as used herein refers to any new and abnormal growth of tissue. Thus, a neoplasm can be a premalignant neoplasm or a malignant neoplasm.

The term "neoplasm-specific marker," as used herein, refers to any biological material or element that can be used to indicate the presence of a neoplasm. Examples of biological materials include, without limitation, nucleic acids, polypeptides, carbohydrates, fatty acids,

cellular components (*e.g.*, cell membranes and mitochondria), and whole cells. In some instances, markers are particular nucleic acid regions (*e.g.*, genes, intragenic regions, specific loci, etc.). Regions of nucleic acid that are markers may be referred to, *e.g.*, as "marker genes," "marker regions," "marker sequences," "marker loci," etc.

As used herein, the term “adenoma” refers to a benign tumor of glandular origin. Although these growths are benign, over time they may progress to become malignant.

The term “pre-cancerous” or “pre-neoplastic” and equivalents thereof refer to any cellular proliferative disorder that is undergoing malignant transformation.

A “site” of a neoplasm, adenoma, cancer, etc. is the tissue, organ, cell type, anatomical area, body part, etc. in a subject’s body where the neoplasm, adenoma, cancer, etc. is located.

As used herein, a “diagnostic” test application includes the detection or identification of a disease state or condition of a subject, determining the likelihood that a subject will contract a given disease or condition, determining the likelihood that a subject with a disease or condition will respond to therapy, determining the prognosis of a subject with a disease or condition (or its likely progression or regression), and determining the effect of a treatment on a subject with a disease or condition. For example, a diagnostic can be used for detecting the presence or likelihood of a subject contracting a neoplasm or the likelihood that such a subject will respond favorably to a compound (e.g., a pharmaceutical, e.g., a drug) or other treatment.

The term “isolated” when used in relation to a nucleic acid, as in “an isolated oligonucleotide” refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids, such as DNA and RNA, are found in the state they exist in nature. Examples of non-isolated nucleic acids include: a given DNA sequence (e.g., a gene) found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, isolated nucleic acid encoding a particular protein includes, by way of example, such nucleic acid in cells ordinarily expressing the protein, where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or

oligonucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide may be single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide may be double-stranded). An isolated nucleic acid may, after isolation from its natural or typical environment, be combined with other nucleic acids or molecules. For example, an isolated nucleic acid may be present in a host cell into which it has been placed, e.g., for heterologous expression.

The term "purified" refers to molecules, either nucleic acid or amino acid sequences that are removed from their natural environment, isolated, or separated. An "isolated nucleic acid sequence" may therefore be a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated. As used herein, the terms "purified" or "to purify" also refer to the removal of contaminants from a sample. The removal of contaminating proteins results in an increase in the percent of polypeptide or nucleic acid of interest in the sample. In another example, recombinant polypeptides are expressed in plant, bacterial, yeast, or mammalian host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

The term "composition comprising" a given polynucleotide sequence or polypeptide refers broadly to any composition containing the given polynucleotide sequence or polypeptide. The composition may comprise an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The term "sample" is used in its broadest sense. In one sense it can refer to an animal cell or tissue. In another sense, it refers to a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. Environmental samples include environmental material such as surface matter, soil, water, and industrial samples. These examples are not to be construed as limiting the sample types applicable to the present invention.

As used herein, a "remote sample" as used in some contexts relates to a sample indirectly collected from a site that is not the cell, tissue, or organ source of the sample. For

instance, when sample material originating from the cervix is assessed in a stool sample the sample is a remote sample.

As used herein, the terms "patient" or "subject" refer to organisms to be subject to various tests provided by the technology. The term "subject" includes animals, preferably mammals, including humans. In a preferred embodiment, the subject is a primate. In an even more preferred embodiment, the subject is a human. Further with respect to diagnostic methods, a preferred subject is a vertebrate subject. A preferred vertebrate is warm-blooded; a preferred warm-blooded vertebrate is a mammal. A preferred mammal is most preferably a human. As used herein, the term "subject' includes both human and animal subjects. Thus, veterinary therapeutic uses are provided herein. As such, the present technology provides for the diagnosis of mammals such as humans, as well as those mammals of importance due to being endangered, such as Siberian tigers; of economic importance, such as animals raised on farms for consumption by humans; and/or animals of social importance to humans, such as animals kept as pets or in zoos. Examples of such animals include but are not limited to: carnivores such as cats and dogs; swine, including pigs, hogs, and wild boars; ruminants and/or ungulates such as cattle, oxen, sheep, giraffes, deer, goats, bison, and camels; pinnipeds; and horses. Thus, also provided is the diagnosis and treatment of livestock, including, but not limited to, domesticated swine, ruminants, ungulates, horses (including race horses), and the like. The presently-disclosed subject matter further includes a system for diagnosing a cervical cancer, cervical cancer subtype (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia) in a subject. The system can be provided, for example, as a commercial kit that can be used to screen for a risk of cervical cancer, cervical cancer subtype (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia), or diagnose a cervical cancer, cervical cancer subtype, and/or a cervical pre-cancer in a subject from whom a biological sample has been collected. An exemplary system provided in accordance with the present technology includes assessing the methylation state of a marker described herein.

As used herein, the term "kit" refers to any delivery system for delivering materials. In the context of reaction assays, such delivery systems include systems that allow for the storage, transport, or delivery of reaction reagents (e.g., oligonucleotides, enzymes, etc. in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions for

performing the assay etc.) from one location to another. For example, kits include one or more enclosures (e.g., boxes) containing the relevant reaction reagents and/or supporting materials. As used herein, the term "fragmented kit" refers to delivery systems comprising two or more separate containers that each contain a subportion of the total kit components. The containers may be delivered to the intended recipient together or separately. For example, a first container may contain an enzyme for use in an assay, while a second container contains oligonucleotides. The term "fragmented kit" is intended to encompass kits containing Analyte specific reagents (ASR's) regulated under section 520(e) of the Federal Food, Drug, and Cosmetic Act, but are not limited thereto. Indeed, any delivery system comprising two or more separate containers that each contains a subportion of the total kit components are included in the term "fragmented kit." In contrast, a "combined kit" refers to a delivery system containing all of the components of a reaction assay in a single container (e.g., in a single box housing each of the desired components). The term "kit" includes both fragmented and combined kits.

As used herein, the term "information" refers to any collection of facts or data. In reference to information stored or processed using a computer system(s), including but not limited to internets, the term refers to any data stored in any format (*e.g.*, analog, digital, optical, *etc.*). As used herein, the term "information related to a subject" refers to facts or data pertaining to a subject (*e.g.*, a human, plant, or animal). The term "genomic information" refers to information pertaining to a genome including, but not limited to, nucleic acid sequences, genes, percentage methylation, allele frequencies, RNA expression levels, protein expression, phenotypes correlating to genotypes, *etc.* "Allele frequency information" refers to facts or data pertaining to allele frequencies, including, but not limited to, allele identities, statistical correlations between the presence of an allele and a characteristic of a subject (*e.g.*, a human subject), the presence or absence of an allele in an individual or population, the percentage likelihood of an allele being present in an individual having one or more particular characteristics, *etc.*

DETAILED DESCRIPTION

Provided herein is technology for cervical cancer screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of cervical cancer and/or specific forms of cervical cancer (e.g., cervical adenocarcinoma, squamous cell cervical cancer, cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia), or

for discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers) from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample).

Indeed, as described in Examples I and II, experiments conducted during the course for identifying embodiments for the present invention identified a novel set of differentially methylated regions (DMRs) for discriminating cancer of the cervix derived DNA from non-neoplastic control DNA, and for discriminating cervical cancer tissue from endometrial and ovarian cancer tissue. Such experiments list and describe 423 novel DNA methylation markers distinguishing cervical cancer, cervical cancer subtypes, and cervical pre-cancers tissue from benign cervical tissue (see, Tables I-IV, VI-VIII, Example I), and cervical cancer tissue from endometrial cancer tissue and ovarian cancer tissue (see, Tables XI and XII, Example II).

In particular aspects, the present technology provides compositions and methods for identifying, determining, and/or classifying a cancer such as cervical cancer, a subtype of cervical cancer (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia). The methods comprise determining the methylation status of at least one methylation marker in a biological sample isolated from a subject (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample), wherein a change in the methylation state of the marker is indicative of the presence, class, or site of cervical cancer, a subtype of cervical cancer (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia). Particular embodiments relate to markers comprising a differentially methylated region (DMR, e.g., DMR 1–423, see Tables I, III, and X) that are used for diagnosis (e.g., screening) of cervical cancer, a subtype of cervical cancer, and/or a cervical pre-cancer, or discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers).

In certain embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA) in a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) obtained from a subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated dinucleotides (e.g., CpG dinucleotides) within one or more methylation markers; and

2) detecting cervical cancer, a cervical cancer subtype (e.g., cervical adenocarcinoma, squamous cell cervical cancer), a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia), or discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers) (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In certain embodiments of the technology, methods are provided that comprise the following steps:

1) measuring a methylation level for one or more genes or methylation markers in biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) of a human individual through treating genomic DNA in the biological sample with a reagent that modifies DNA in a methylation-specific manner;

2) amplifying the treated genomic DNA using a set of primers for the selected one or more genes or methylated markers; and

3) determining the methylation level of the one or more genes or methylated markers.

In certain embodiments of the technology, methods are provided that comprise the following steps:

1) measuring an amount of one or more methylated marker genes in DNA from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample);

2) measuring the amount of at least one reference marker in the DNA; and

3) calculating a value for the amount of the one or more methylated marker genes measured in the DNA as a percentage of the amount of the reference marker gene measured in the DNA, wherein the value indicates the amount of the one or more methylated marker DNA measured in the biological sample.

In certain embodiments of the technology, methods are provided that comprise the following steps:

1) measuring a methylation level of a CpG site for one or more genes in a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) of a human individual through treating genomic DNA in the biological sample with bisulfite a reagent capable of modifying DNA in a methylation-specific manner;

2) amplifying the modified genomic DNA using a set of primers for the selected one or more genes; and

3) determining the methylation level of the CpG site for the selected one or more genes.

In certain embodiments, the technology provides methods for characterizing a biological sample comprising:

(a) measuring a methylation level of a CpG site for one or more genes in a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) of a human individual through treating genomic DNA in the biological sample with bisulfite; amplifying the bisulfite-treated genomic DNA using a set of primers for the selected one or more genes; and determining the methylation level of the CpG site; and

(b) comparing the methylation level of the one or more genes to a methylation level of a corresponding set of genes in control samples without cervical cancer; and

(c) determining that the individual has cervical cancer, a subtype of cervical cancer (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia) when the methylation level measured in the one or more genes is higher than the methylation level measured in the respective control samples.

In certain embodiments, the technology provides methods for discriminating cervical cancer from other types of gynecological cancers (e.g., endometrial, and ovarian cancers) in biological sample comprising:

(a) measuring a methylation level of a CpG site for one or more genes in a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) of a human individual through treating genomic DNA in the biological sample with bisulfite; amplifying the bisulfite-treated genomic DNA using a set of primers for the selected one or more genes; and determining the methylation level of the CpG site; and

(b) comparing the methylation level of the one or more genes to a methylation level of a corresponding set of genes in endometrial cancer samples and/or ovarian cancer samples; and

(c) determining that the individual has cervical cancer when the methylation level measured in the one or more genes is higher than the methylation level measured in the respective endometrial cancer and/or ovarian cancer samples.

In certain embodiments, the technology provides methods of screening for cervical cancer, a subtype of cervical cancer (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia) in a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) obtained from a subject, the method comprising

1) assaying a methylation state of one or more DNA methylation markers; and

2) identifying the subject as having cervical cancer, a subtype of cervical cancer (e.g., cervical adenocarcinoma, squamous cell cervical cancer), and/or a cervical pre-cancer (e.g., cervix related in-situ adenocarcinoma, cervical intraepithelial neoplasia) when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have cervical cancer.

In certain embodiments, the technology provides methods for characterizing a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion,

vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) comprising measuring an amount of one or more methylated markers gene in DNA extracted from the biological sample; treating genomic DNA in the biological sample with bisulfite; amplifying the bisulfite-treated genomic DNA using primers specific for a CpG site for each marker gene, wherein the primers specific for each marker gene are capable of binding an amplicon bound by a primer sequence for the marker gene (e.g., a primer recited in Tables V and/or XII), wherein the amplicon bound by the primer sequence for the marker gene is at least a portion of a genetic region for the methylated marker gene recited in Tables I, III and/or X; determining the methylation level of the CpG site for one or more genes.

In certain embodiments, the technology provides methods comprising measuring the methylation level of one or more methylated marker genes in DNA extracted from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) through extracting genomic DNA from a biological sample of a human individual suspected of having or having cancer; treating the extracted genomic DNA with bisulfite, amplifying the bisulfite-treated genomic DNA with primers specific for the one or more genes, wherein the primers specific for the one or more genes are capable of binding at least a portion of the bisulfite-treated genomic DNA for a chromosomal region for the marker recited in Tables I, III, and X; and measuring the methylation level of one or more methylated marker genes.

In certain embodiments, the technology provides methods for preparing a DNA fraction from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) of a human individual useful for analyzing one or more genetic loci involved in one or more chromosomal aberrations, comprising:

(a) extracting genomic DNA from a biological sample of a human individual;

(b) producing a fraction of the extracted genomic DNA by:

(i) treating the extracted genomic DNA with a reagent that modifies DNA in a methylation-specific manner;

(ii) amplifying the bisulfite-treated genomic DNA using separate primers specific for one or more methylation markers;

(c) analyzing one or more genetic loci in the produced fraction of the extracted genomic DNA by measuring a methylation level of the CpG site for each of the one or more methylation markers.

In certain embodiments, the technology provides methods for preparing a DNA fraction from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) of a human individual useful for analyzing one or more DNA fragments involved in one or more chromosomal aberrations, comprising:

(a) extracting genomic DNA from a biological sample of a human individual;

(b) producing a fraction of the extracted genomic DNA by:

(i) treating the extracted genomic DNA with a reagent that modifies DNA in a methylation-specific manner;

(ii) amplifying the bisulfite-treated genomic DNA using separate primers specific for one or more methylation markers; and

(c) analyzing one or more DNA fragments in the produced fraction of the extracted genomic DNA by measuring a methylation level of the CpG site for each of the one or more methylation markers.

Preferably, the sensitivity for such methods is from about 70% to about 100%, or from about 80% to about 90%, or from about 80% to about 85%. Preferably, the specificity is from about 70% to about 100%, or from about 80% to about 90%, or from about 80% to about 85%.

Such methods are not limited to specific methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers. In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers comprise a base in a DMR selected from a group consisting of DMR 1–423 as provided in Tables I, III, and X.

In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers are selected from Tables I and/or III.

In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers are selected from MAX.chr6.58147682-

58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5 (Example I).

In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers are selected from C1orf114, MAX.chr6.58147682-58147771, ZNF773, NEUROG3, ASCL1, NID2, ZNF781, CRHR2, and MAX.chr9.36739811-36739868 (see, Table VI and Example I).

In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers are selected from ABCB1, ARHGAP12, ASCL1, BARHL1, C1orf114, C2orf40, CACNA1C, CRHR2, HOPX_C, KCNQ5, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, PRDM12, SLC9A3, TMEM200C, TTYH1, ZNF382, ZNF773, and ZNF781 (see, Table VII, Example I).

In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers are selected from ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18 (see, Table VIII, Example I).

In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers are selected from Table X.

In some embodiments, the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers are selected from AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1 (see, Tables X and XI, Example II).

Such methods are not limited to a specific sample or biological sample type. For example, in some embodiments the biological sample is a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample..

Various cancers (e.g., cervical cancer and cervical cancer subtypes) and pre-cancers (e.g., cervical pre-cancers) are predicted by various combinations of markers, e.g., as identified by statistical techniques related to specificity and sensitivity of prediction. The technology further provides methods for identifying predictive combinations and validated predictive combinations for some cancers.

Such methods are not limited to a subject type. In some embodiments, the subject is a mammal. In some embodiments, the subject is a human.

Such methods are not limited to a particular manner or technique for determining characterizing, measuring, or assaying methylation for one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers. In some embodiments, such techniques are based upon an analysis of the methylation status (e.g., CpG methylation status) of at least one marker, region of a marker, or base of a marker comprising a DMR.

In some embodiments, measuring the methylation state of a methylation marker in a sample comprises determining the methylation state of one base. In some embodiments, measuring the methylation state of the marker in the sample comprises determining the extent of methylation at a plurality of bases. Moreover, in some embodiments, the methylation state of the methylated marker comprises an increase in methylation of the marker relative to a normal methylation state of the marker. In some embodiments, the methylation state of the marker comprises a decreased methylation of the marker relative to a normal methylation state of the marker. In some embodiments the methylation state of the marker comprises a different pattern of methylation of the marker relative to a normal methylation state of the marker.

Furthermore, in some embodiments the marker is a region of 100 or fewer bases, the marker is a region of 500 or fewer bases, the marker is a region of 1000 or fewer bases, the marker is a region of 5000 or fewer bases, or, in some embodiments, the marker is one base. In some embodiments the marker is in a high CpG density promoter.

In certain embodiments, methods for analyzing a nucleic acid for the presence of 5-methylcytosine involves treatment of DNA with a reagent that modifies DNA in a methylation-specific manner. Examples of such reagents include, but are not limited to, a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, a bisulfite reagent, a TET enzyme, and a borane reducing agent.

A frequently used method for analyzing a nucleic acid for the presence of 5-methylcytosine is based upon the bisulfite method described by Frommer, et al. for the detection of 5-methylcytosines in DNA (Frommer et al. (1992) *Proc. Natl. Acad. Sci. USA* 89: 1827–31 explicitly incorporated herein by reference in its entirety for all purposes) or variations thereof. The bisulfite method of mapping 5-methylcytosines is based on the observation that cytosine, but not 5-methylcytosine, reacts with hydrogen sulfite ion (also known as bisulfite). The reaction is usually performed according to the following steps: first, cytosine reacts with hydrogen sulfite to form a sulfonated cytosine. Next, spontaneous deamination of the sulfonated reaction intermediate results in a sulfonated uracil. Finally, the sulfonated uracil is desulfonated under alkaline conditions to form uracil. Detection is possible because uracil base pairs with adenine (thus behaving like thymine), whereas 5-methylcytosine base pairs with guanine (thus behaving like cytosine). This makes the discrimination of methylated cytosines from non-methylated cytosines possible by, *e.g.*, bisulfite genomic sequencing (Grigg G, & Clark S, Bioessays (1994) 16: 431–36; Grigg G, DNA Seq. (1996) 6: 189–98),methylation-specific PCR (MSP) as is disclosed, *e.g.*, in U.S. Patent No. 5,786,146, or using an assay comprising sequence-specific probe cleavage, *e.g.*, a QuARTS flap endonuclease assay (see, *e.g.*, Zou et al. (2010) "Sensitive quantification of methylated markers with a novel methylation specific technology" *Clin Chem* **56**: A199; and in U.S. Pat. Nos. 8,361,720; 8,715,937; 8,916,344; and 9,212,392.

Some conventional technologies are related to methods comprising enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing precipitation and purification steps with a fast dialysis (Olek A, et al. (1996) "A modified and improved method for bisulfite based cytosine methylation analysis" *Nucleic Acids Res.* 24: 5064-6). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of conventional methods for detecting 5-methylcytosine is provided by Rein, T., et al. (1998) *Nucleic Acids Res.* 26: 2255.

The bisulfite technique typically involves amplifying short, specific fragments of a known nucleic acid subsequent to a bisulfite treatment, then either assaying the product by sequencing (Olek & Walter (1997) *Nat. Genet.* 17: 275–6) or a primer extension reaction (Gonzalgo & Jones (1997) *Nucleic Acids Res.* 25: 2529–31; WO 95/00669; U.S. Pat. No. 6,251,594) to analyze individual cytosine positions. Some methods use enzymatic digestion (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532–4). Detection by hybridization has also

been described in the art (Olek et al., WO 99/28498). Additionally, use of the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg & Clark (1994) *Bioessays* 16: 431–6,; Zeschnigk et al. (1997) *Hum Mol Genet.* 6: 387–95; Feil et al. (1994) *Nucleic Acids Res.* 22: 695; Martin et al. (1995) *Gene* 157: 261–4; WO 9746705; WO 9515373).

Various methylation assay procedures can be used in conjunction with bisulfite treatment according to the present technology. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (*e.g.*, CpG islands) within a nucleic acid sequence. Such assays involve, among other techniques, sequencing of bisulfite-treated nucleic acid, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-specific restriction enzymes, *e.g.*, methylation-sensitive or methylation-dependent enzymes.

For example, genomic sequencing has been simplified for analysis of methylation patterns and 5-methylcytosine distributions by using bisulfite treatment (Frommer et al. (1992) *Proc. Natl. Acad. Sci. USA* 89: 1827–1831). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA finds use in assessing methylation state, e.g., as described by Sadri & Hornsby (1997) *Nucl. Acids Res.* 24: 5058–5059 or as embodied in the method known as COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532–2534).

COBRA™ analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific loci in small amounts of genomic DNA (Xiong & Laird, Nucleic Acids Res. 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (Proc. Natl. Acad. Sci. USA 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the CpG islands of interest, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples.

Typical reagents (e.g., as might be found in a typical COBRA™-based kit) for COBRA™ analysis may include, but are not limited to: PCR primers for specific loci (e.g., specific genes, markers, DMR, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, etc.); restriction enzyme and appropriate buffer; gene-hybridization oligonucleotide; control hybridization oligonucleotide; kinase labeling kit for oligonucleotide probe; and labeled nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components. Assays such as "MethyLight™" (a fluorescence-based real-time PCR technique) (Eads et al., Cancer Res. 59:2302-2306, 1999), Ms-SNuPE™ (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; U.S. Pat. No. 5,786,146), and methylated CpG island amplification ("MCA"; Toyota et al., Cancer Res. 59:2307-12, 1999) are used alone or in combination with one or more of these methods.

The "HeavyMethyl™" assay, technique is a quantitative method for assessing methylation differences based on methylation-specific amplification of bisulfite-treated DNA. Methylation-specific blocking probes ("blockers") covering CpG positions between, or covered by, the amplification primers enable methylation-specific selective amplification of a nucleic acid sample.

The term "HeavyMethyl™ MethyLight™" assay refers to a HeavyMethyl™ MethyLight™ assay, which is a variation of the MethyLight™ assay, wherein the MethyLight™ assay is combined with methylation specific blocking probes covering CpG positions between the amplification primers. The HeavyMethyl™ assay may also be used in combination with methylation specific amplification primers.

Typical reagents (*e.g.*, as might be found in a typical MethyLight™-based kit) for HeavyMethyl™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, or bisulfite treated DNA sequence or CpG island, *etc.*); blocking oligonucleotides; optimized PCR buffers and deoxynucleotides; and Taq polymerase. MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; U.S. Pat.

No. 5,786,146). Briefly, DNA is modified by sodium bisulfite, which converts unmethylated, but not methylated cytosines, to uracil, and the products are subsequently amplified with primers specific for methylated versus unmethylated DNA. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); optimized PCR buffers and deoxynucleotides, and specific probes.

The MethyLight™ assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (*e.g.*, TaqMan®) that requires no further manipulations after the PCR step (Eads et al., Cancer Res. 59:2302-2306, 1999). Briefly, the MethyLight™ process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed in a "biased" reaction, *e.g.*, with PCR primers that overlap known CpG dinucleotides. Sequence discrimination occurs both at the level of the amplification process and at the level of the fluorescence detection process.

The MethyLight™ assay is used as a quantitative test for methylation patterns in a nucleic acid, *e.g.*, a genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In a quantitative version, the PCR reaction provides for a methylation specific amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe, overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing the biased PCR pool with either control oligonucleotides that do not cover known methylation sites (*e.g.*, a fluorescence-based version of the HeavyMethyl™ and MSP techniques) or with oligonucleotides covering potential methylation sites.

The MethyLight™ process is used with any suitable probe (*e.g.* a "TaqMan®" probe, a Lightcycler® probe, *etc.*) For example, in some applications double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes, *e.g.*, with MSP primers and/or HeavyMethyl blocker oligonucleotides and

a TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules and is designed to be specific for a relatively high GC content region so that it melts at about a 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5′ to 3′ endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Typical reagents (*e.g.*, as might be found in a typical MethyLight™-based kit) for MethyLight™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); TaqMan® or Lightcycler® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

The QM™ (quantitative methylation) assay is an alternative quantitative test for methylation patterns in genomic DNA samples, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe, overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing the biased PCR pool with either control oligonucleotides that do not cover known methylation sites (a fluorescence-based version of the HeavyMethyl™ and MSP techniques) or with oligonucleotides covering potential methylation sites.

The QM™ process can be used with any suitable probe, *e.g.*, "TaqMan®" probes, Lightcycler® probes, in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to unbiased primers and the TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about a 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5′ to

3′ endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system. Typical reagents (*e.g.*, as might be found in a typical QM™-based kit) for QM™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); TaqMan® or Lightcycler® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

The Ms-SNuPE™ technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site of interest. Small amounts of DNA can be analyzed (*e.g.*, microdissected pathology sections) and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

Typical reagents (*e.g.*, as might be found in a typical Ms-SNuPE™-based kit) for Ms-SNuPE™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE™ primers for specific loci; reaction buffer (for the Ms-SNuPE reaction); and labeled nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Reduced Representation Bisulfite Sequencing (RRBS) begins with bisulfite treatment of nucleic acid to convert all unmethylated cytosines to uracil, followed by restriction enzyme digestion (*e.g.*, by an enzyme that recognizes a site including a CG sequence such as MspI) and complete sequencing of fragments after coupling to an adapter ligand. The choice of restriction enzyme enriches the fragments for CpG dense regions, reducing the number of redundant sequences that may map to multiple gene positions during analysis. As such, RRBS reduces the complexity of the nucleic acid sample by selecting a subset (*e.g.*, by size

selection using preparative gel electrophoresis) of restriction fragments for sequencing. As opposed to whole-genome bisulfite sequencing, every fragment produced by the restriction enzyme digestion contains DNA methylation information for at least one CpG dinucleotide. As such, RRBS enriches the sample for promoters, CpG islands, and other genomic features with a high frequency of restriction enzyme cut sites in these regions and thus provides an assay to assess the methylation state of one or more genomic loci.

A typical protocol for RRBS comprises the steps of digesting a nucleic acid sample with a restriction enzyme such as MspI, filling in overhangs and A-tailing, ligating adaptors, bisulfite conversion, and PCR. See, *e.g.*, et al. (2005) "Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution" *Nat Methods* **7**: 133–6; Meissner et al. (2005) "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis" *Nucleic Acids Res.* **33**: 5868–77.

In some embodiments, a quantitative allele-specific real-time target and signal amplification (QuARTS) assay is used to evaluate methylation state. Three reactions sequentially occur in each QuARTS assay, including amplification (reaction 1) and target probe cleavage (reaction 2) in the primary reaction; and FRET cleavage and fluorescent signal generation (reaction 3) in the secondary reaction. When target nucleic acid is amplified with specific primers, a specific detection probe with a flap sequence loosely binds to the amplicon. The presence of the specific invasive oligonucleotide at the target binding site causes a 5′ nuclease, *e.g.*, a FEN-1 endonuclease, to release the flap sequence by cutting between the detection probe and the flap sequence. The flap sequence is complementary to a non-hairpin portion of a corresponding FRET cassette. Accordingly, the flap sequence functions as an invasive oligonucleotide on the FRET cassette and effects a cleavage between the FRET cassette fluorophore and a quencher, which produces a fluorescent signal. The cleavage reaction can cut multiple probes per target and thus release multiple fluorophores per flap, providing exponential signal amplification. QuARTS can detect multiple targets in a single reaction well by using FRET cassettes with different dyes. See, *e.g.*, in Zou et al. (2010) "Sensitive quantification of methylated markers with a novel methylation specific technology" *Clin Chem* **56**: A199), and U.S. Pat. Nos. 8,361,720; 8,715,937; 8,916,344; and 9,212,392, each of which is incorporated herein by reference for all purposes.

The term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite, or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences. Methods of said treatment are

known in the art (*e.g.*, PCT/EP2004/011715 and WO 2013/116375, each of which is incorporated by reference in its entirety). In some embodiments, bisulfite treatment is conducted in the presence of denaturing solvents such as but not limited to n-alkyleneglycol or diethylene glycol dimethyl ether (DME), or in the presence of dioxane or dioxane derivatives. In some embodiments the denaturing solvents are used in concentrations between 1% and 35% (v/v). In some embodiments, the bisulfite reaction is carried out in the presence of scavengers such as but not limited to chromane derivatives, *e.g.*, 6-hydroxy-2,5,7,8,-tetramethylchromane 2-carboxylic acid or trihydroxybenzone acid and derivates thereof, *e.g.*, Gallic acid (see: PCT/EP2004/011715, which is incorporated by reference in its entirety). In certain preferred embodiments, the bisulfite reaction comprises treatment with ammonium hydrogen sulfite, *e.g.*, as described in WO 2013/116375.

In some embodiments, fragments of the treated DNA are amplified using sets of primer oligonucleotides according to the present invention (e.g., see Tables V and XII) and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). Amplicons are typically 100 to 2000 base pairs in length.

In another embodiment of the method, the methylation status of CpG positions within or near a marker comprising a DMR (e.g., DMR 1–423, Tables I, III and X) may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in U.S. Pat. No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primer pairs contain at least one primer that hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the position of the C position in the CpG.

Such methods are not limited to a specific type or kind of primer or primer pair related to the one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers. In some embodiments, the primer or primer pair is recited in Table V (SEQ ID Nos: 1-76). In some embodiments, the primer or primer pair specific for each methylated marker gene are capable of binding an amplicon bound by a primer sequence for the marker gene recited in Tables V and/or XII, wherein the amplicon bound by the primer sequence for the marker gene recited in Tables V and/or XII is at least a portion of a

genetic region for the methylated marker gene recited in Tables I, III, and X. In some embodiments, the primer or primer pair for a methylated marker is a set of primers that specifically binds at least a portion of a genetic region comprising chromosomal coordinates for the specific methylated marker recited in Tables I, III, and X.

In another embodiment, the invention provides a method for converting an oxidized 5-methylcytosine residue in cell-free DNA to a dihydrouracil residue (see, Liu et al., 2019, Nat Biotechnol. 37, pp. 424-429; U.S. Patent Application Publication No. 202000370114). The method involves reaction of an oxidized 5mC residue selected from 5-formylcytosine (5fC), 5-carboxymethylcytosine (5caC), and combinations thereof, with a borane reducing agent. The oxidized 5mC residue may be naturally occurring or, more typically, the result of a prior oxidation of a 5mC or 5hmC residue, e.g., oxidation of 5mC or 5hmC with a TET family enzyme (e.g., TET1, TET2, or TET3), or chemical oxidation of 5 mC or 5hmC, e.g., with potassium perruthenate ($KRuO_4$) or an inorganic peroxo compound or composition such as peroxotungstate (see, e.g., Okamoto et al. (2011) *Chem. Commun.* 47:11231-33) and a copper (II) perchlorate/2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) combination (see Matsushita et al. (2017) *Chem. Commun.* 53:5756-59).

The borane reducing agent may be characterized as a complex of borane and a nitrogen-containing compound selected from nitrogen heterocycles and tertiary amines. The nitrogen heterocycle may be monocyclic, bicyclic, or polycyclic, but is typically monocyclic, in the form of a 5- or 6-membered ring that contains a nitrogen heteroatom and optionally one or more additional heteroatoms selected from N, O, and S. The nitrogen heterocycle may be aromatic or alicyclic. Preferred nitrogen heterocycles herein include 2-pyrroline, 2H-pyrrole, 1H-pyrrole, pyrazolidine, imidazolidine, 2-pyrazoline, 2-imidazoline, pyrazole, imidazole, 1,2,4-triazole, 1,2,4-triazole, pyridazine, pyrimidine, pyrazine, 1,2,4-triazine, and 1,3,5-triazine, any of which may be unsubstituted or substituted with one or more non-hydrogen substituents. Typical non-hydrogen substituents are alkyl groups, particularly lower alkyl groups, such as methyl, ethyl, n-propyl, isopropyl, n-butyl, isobutyl, t-butyl, and the like. Exemplary compounds include pyridine borane, 2-methylpyridine borane (also referred to as 2-picoline borane), and 5-ethyl-2-pyridine.

The reaction of the borane reducing agent with the oxidized 5mC residue in cell-free DNA is advantageous insofar as non-toxic reagents and mild reaction conditions can be employed; there is no need for any bisulfate, nor for any other potentially DNA-degrading reagents. Furthermore, conversion of an oxidized 5mC residue to dihydrouracil with the

borane reducing agent can be carried out without need for isolation of any intermediates, in a "one-pot" or "one-tube" reaction. This is quite significant, since the conversion involves multiple steps, i.e., (1) reduction of the alkene bond linking C-4 and C-5 in the oxidized 5mC, (2) deamination, and (3) either decarboxylation, if the oxidized 5mC is 5caC, or deformylation, if the oxidized 5mC is 5fC.

In addition to a method for converting an oxidized 5-methylcytosine residue in cell-free DNA to a dihydrouracil residue, the invention also provides a reaction mixture related to the aforementioned method. The reaction mixture comprises a sample of cell-free DNA containing at least one oxidized 5-methylcytosine residue selected from 5caC, 5fC, and combinations thereof, and a borane reducing agent effective to effective to reduce, deaminate, and either decarboxylate or deformylate the at least one oxidized 5-methylcytosine residue. The borane reducing agent is a complex of borane and a nitrogen-containing compound selected from nitrogen heterocycles and tertiary amines, as explained above. In a preferred embodiment, the reaction mixture is substantially free of bisulfite, meaning substantially free of bisulfite ion and bisulfite salts. Ideally, the reaction mixture contains no bisulfite.

In a related aspect of the invention, a kit is provided for converting 5mC residues in cell-free DNA to dihydrouracil residues, where the kit includes a reagent for blocking 5hmC residues, a reagent for oxidizing 5mC residues beyond hydroxymethylation to provide oxidized 5mC residues, and a borane reducing agent effective to reduce, deaminate, and either decarboxylate or deformylate the oxidized 5mC residues. The kit may also include instructions for using the components to carry out the above-described method.

In another embodiment, a method is provided that makes use of the above-described oxidation reaction. The method enables detecting the presence and location of 5-methylcytosine residues in cell-free DNA, and comprises the following steps:

(a) modifying 5hmC residues in fragmented, adapter-ligated cell-free DNA to provide an affinity tag thereon, wherein the affinity tag enables removal of modified 5hmC-containing DNA from the cell-free DNA;

(b) removing the modified 5hmC-containing DNA from the cell-free DNA, leaving DNA containing unmodified 5mC residues;

(c) oxidizing the unmodified 5mC residues to give DNA containing oxidized 5mC residues selected from 5caC, 5fC, and combinations thereof;

(d) contacting the DNA containing oxidized 5mC residues with a borane reducing agent effective to reduce, deaminate, and either decarboxylate or deformylate the oxidized

5mC residues, thereby providing DNA containing dihydrouracil residues in place of the oxidized 5mC residues;

(e) amplifying and sequencing the DNA containing dihydrouracil residues;

(f) determining a 5-methylation pattern from the sequencing results in (e).

In another embodiments, a method is provided for identifying 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC) in a target nucleic acid comprising the steps of:

providing a biological sample comprising the target nucleic acid;

modifying the target nucleic acid comprising the steps of:

converting the 5mC and 5hmC in the nucleic acid sample to 5-carboxylcytosine (5caC) and/or 5-formylcytosine (5fC) by contacting the nucleic acid sample with a TET enzyme so that one or more 5caC or 5fC residues are generated; and

converting the 5caC and/or 5fC to dihydrouracil (DHU) by treating the target nucleic acid with a borane reducing agent to provide a modified nucleic acid sample comprising a modified target nucleic acid; and

detecting the sequence of the modified target nucleic acid; wherein a cytosine (C) to thymine (T) transition or a cytosine (C) to DHU transition in the sequence of the modified target nucleic acid compared to the target nucleic acid provides the location of either a 5mC or 5hmC in the target nucleic acid.

In some embodiments, the borane reducing agent is 2-picoline borane.

In some embodiments, the step of detecting the sequence of the modified target nucleic acid comprises one or more of chain termination sequencing, microarray, high-throughput sequencing, and restriction enzyme analysis.

In some embodiments, the TET enzyme is selected from the group consisting of human TET1, TET2, and TET3; murine Tet1, Tet2, and Tet3; Naegleria TET (NgTET); and Coprinopsis cinerea (CcTET).

In some embodiments, the method further comprises a step of blocking one or more modified cytosines. In some embodiments, the step of blocking comprises adding a sugar to a 5hmC.

In some embodiments, the method further comprises a step of amplifying the copy number of one or more nucleic acid sequences.

In some embodiments, the oxidizing agent is potassium perruthenate or Cu(II)/TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl.)

The cell-free DNA is extracted from a body sample from a subject, where the body sample is typically whole blood, plasma, or serum, most typically plasma, but the sample may also be tissue (e.g., cervical tissue), a secretion (e.g., cervical secretion, vaginal secretion), an organ secretion, CSF, urine, saliva, mucosal excretions, sputum, stool, or tears. In some embodiments, the cell-free DNA is derived from a tumor. In other embodiments, the cell-free DNA is from a patient with a disease or other pathogenic condition. The cell-free DNA may or may not derive from a tumor. In step (a), it should be noted that the cell-free DNA in which 5hmC residues are to be modified is in purified, fragmented form, and adapter-ligated. DNA purification in this context can be carried out using any suitable method known to those of ordinary skill in the art and/or described in the pertinent literature, and, while cell-free DNA can itself be highly fragmented, further fragmentation may occasionally be desirable, as described, for example, in U.S. Patent Publication No. 2017/0253924. The cell-free DNA fragments are generally in the size range of about 20 nucleotides to about 500 nucleotides, more typically in the range of about 20 nucleotides to about 250 nucleotides. The purified cell-free DNA fragments that are modified in step (a) have been end-repaired using conventional means (e.g., a restriction enzyme) so that the fragments have a blunt end at each 3′ and 5′ terminus. In a preferred method, as described in WO 2017/176630, the blunted fragments have also been provided provided with a 3′ overhang comprising a single adenine residue using a polymerase such as Taq polymerase. This facilitates subsequent ligation of a selected universal adapter, i.e., an adapter such as a Y-adapter or a hairpin adapter that ligates to both ends of the cell-free DNA fragments and contains at least one molecular barcode. Use of adapters also enables selective PCR enrichment of adapter-ligated DNA fragments.

In step (a), then, the "purified, fragmented cell-free DNA" comprises adapter-ligated DNA fragments. Modification of 5hmC residues in these cell-free DNA fragments with an affinity tag, as specified in step (a), is done so as to enable subsequent removal of the modified 5hmC-containing DNA from the cell-free DNA. In one embodiment, the affinity tag comprises a biotin moiety, such as biotin, desthiobiotin, oxybiotin, 2-iminobiotin, diaminobiotin, biotin sulfoxide, biocytin, or the like. Use of a biotin moiety as the affinity tag allows for facile removal with streptavidin, e.g., streptavidin beads, magnetic streptavidin beads, etc.

Tagging 5hmC residues with a biotin moiety or other affinity tag is accomplished by covalent attachment of a chemoselective group to 5hmC residues in the DNA fragments, where the chemoselective group is capable of undergoing reaction with a functionalized

affinity tag so as to link the affinity tag to the 5hmC residues. In one embodiment, the chemoselective group is UDP glucose-6-azide, which undergoes a spontaneous 1,3-cycloaddition reaction with an alkyne-functionalized biotin moiety, as described in Robertson et al. (2011) *Biochem. Biophys. Res. Comm.* 411(1):40-3, U.S. Pat. No. 8,741,567, and WO 2017/176630. Addition of an alkyne-functionalized biotin-moiety thus results in covalent attachment of the biotin moiety to each 5hmC residue.

The affinity-tagged DNA fragments can then be pulled down in step (b) using, in one embodiment, streptavidin, in the form of streptavidin beads, magnetic streptavidin beads, or the like, and set aside for later analysis, if so desired. The supernatant remaining after removal of the affinity-tagged fragments contains DNA with unmodified 5mC residues and no 5hmC residues.

In step (c), the unmodified 5mC residues are oxidized to provide 5caC residues and/or 5fC residues, using any suitable means. The oxidizing agent is selected to oxidize 5mC residues beyond hydroxymethylation, i.e., to provide 5caC and/or 5fC residues. Oxidation may be carried out enzymatically, using a catalytically active TET family enzyme. A "TET family enzyme" or a "TET enzyme" as those terms are used herein refer to a catalytically active "TET family protein" or a "TET catalytically active fragment" as defined in U.S. Pat. No. 9,115,386, the disclosure of which is incorporated by reference herein. A preferred TET enzyme in this context is TET2; see Ito et al. (2011) *Science* 333(6047):1300-1303. Oxidation may also be carried out chemically, as described in the preceding section, using a chemical oxidizing agent. Examples of suitable oxidizing agent include, without limitation: a perruthenate anion in the form of an inorganic or organic perruthenate salt, including metal perruthenates such as potassium perruthenate ($KRuO_4$), tetraalkylammonium perruthenates such as tetrapropylammonium perruthenate (TPAP) and tetrabutylammonium perruthenate (TBAP), and polymer supported perruthenate (PSP); and inorganic peroxo compounds and compositions such as peroxotungstate or a copper (II) perchlorate/TEMPO combination. It is unnecessary at this point to separate 5fC-containing fragments from 5caC-containing fragments, insofar as in the next step of the process, step (e) converts both 5fC residues and 5caC residues to dihydrouracil (DHU).

In some embodiments, 5-hydroxymethylcytosine residues are blocked with β-glucosyltransferase (β3GT), while 5-methylcytosine residues are oxidized with a TET enzyme effective to provide a mixture of 5-formylcytosine and 5-carboxymethylcytosine. The mixture containing both of these oxidized species can be reacted with 2-picoline borane

or another borane reducing agent to give dihydrouracil. In a variation on this embodiment, 5hmC-containing fragments are not removed in step (b). Rather, "TET-Assisted Picoline Borane Sequencing (TAPS)," 5mC-containing fragments and 5hmC-containing fragments are together enzymatically oxidized to provide 5fC- and 5caC-containing fragments. Reaction with 2-picoline borane results in DHU residues wherever 5mC and 5hmC residues were originally present. "Chemical Assisted Picoline Borane Sequencing (CAPS)," involves selective oxidation of 5hmC-containing fragments with potassium perruthenate, leaving 5mC residues unchanged.

There are numerous advantages to the method of this embodiment: bisulfite is unnecessary, nontoxic reagents and reactants are employed; and the process proceeds under mild conditions. In addition, the entire process can be performed in a single tube, without need for isolation of any intermediates.

In a related embodiment, the above method includes a further step: (g) identifying a hydroxymethylation pattern in the 5hmC-containing DNA removed from the cell-free DNA in step (b). This can be carried out using the techniques described in detail in WO 2017/176630. The process can be carried out without removal or isolation of intermediates in a one-tube method. For example, initially, cell-free DNA fragments, preferably adapter-ligated DNA fragments, are subjected to functionalization with βGT-catalyzed uridine diphosphoglucose 6-azide, followed by biotinylation via the chemoselective azide groups. This procedure results in covalently attached biotin at each 5hmC site. In a next step, the biotinylated strands and strands containing unmodified (native) 5mC are pulled down simultaneously for further processing. The native 5mC-containing strands are pulled down using an anti-5mC antibody or a methyl-CpG-binding domain (MBD) protein, as is known in the art. Then, with the 5hmC residues blocked, the unmodified 5mC residues are selectively oxidized using any suitable technique for converting 5mC to 5fC and/or 5caC, as described elsewhere herein.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. In some embodiments, the labels are fluorescent labels, radionuclides, or detachable molecule fragments having a typical mass that can be detected in a mass spectrometer. Where said labels are mass labels, some embodiments provide that the labeled amplicons have a single positive or negative net charge, allowing for better delectability in the mass spectrometer. The detection may be carried out and visualized by

means of, e.g., matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Methods for isolating DNA suitable for these assay technologies are known in the art. In particular, some embodiments comprise isolation of nucleic acids as described in U.S. Pat. Appl. Ser. No. 13/470,251 ("Isolation of Nucleic Acids"), incorporated herein by reference in its entirety.

In some embodiments, the markers described herein find use in QUARTS assays performed on a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample). In some embodiments, methods for producing DNA samples and, in particular, to methods for producing DNA samples that comprise highly purified, low-abundance nucleic acids in a small volume (*e.g.*, less than 100, less than 60 microliters) and that are substantially and/or effectively free of substances that inhibit assays used to test the DNA samples (*e.g.*, PCR, INVADER, QuARTS assays, *etc.*) are provided. Such DNA samples find use in diagnostic assays that qualitatively detect the presence of, or quantitatively measure the activity, expression, or amount of, a gene, a gene variant (*e.g.*, an allele), or a gene modification (*e.g.*, methylation) present in a sample taken from a patient. For example, some cancers are correlated with the presence of particular mutant alleles or particular methylation states, and thus detecting and/or quantifying such mutant alleles or methylation states has predictive value in the diagnosis and treatment of cancer.

Many valuable genetic markers are present in extremely low amounts in samples and many of the events that produce such markers are rare. Consequently, even sensitive detection methods such as PCR require a large amount of DNA to provide enough of a low-abundance target to meet or supersede the detection threshold of the assay. Moreover, the presence of even low amounts of inhibitory substances compromise the accuracy and precision of these assays directed to detecting such low amounts of a target. Accordingly, provided herein are methods providing the requisite management of volume and concentration to produce such DNA samples.

In some embodiments, the sample comprises a tissue sample (e.g., cervical tissue), blood, plasma, serum, whole blood, a secretion (e.g., cervical secretion, vaginal secretion), an organ secretion, CSF, saliva, urine, or stool. In some embodiments, the subject is human.

Such samples can be obtained by any number of means known in the art, such as will be apparent to the skilled person. For instance, urine and fecal samples are easily attainable, while blood, ascites, serum, or pancreatic fluid samples can be obtained parenterally by using a needle and syringe, for instance. Cell free or substantially cell free samples can be obtained by subjecting the sample to various techniques known to those of skill in the art which include, but are not limited to, centrifugation and filtration. Although it is generally preferred that no invasive techniques are used to obtain the sample, it still may be preferable to obtain samples such as tissue homogenates, tissue sections, and biopsy specimens.

In some embodiments, the sample is obtained with any type or kind of collection device capable of obtaining the desired sample type. For instance, a collection device may be a device capable of obtaining a cervical tissue sample. In certain embodiments, the collection device is a device capable of obtaining tissue or cells from or near the cervix. In some embodiments, a cervical tissue sample includes, for example, a sample comprising any cervical tissue or cervical cells, and may comprise tissue or cells from areas anatomically within the vicinity of the cervix (e.g., vaginal tissue, vaginal cells, endometrial tissue, endometrial cells, ovarian tissue, ovarian cells, etc.) in addition to cervical tissue or cells. In another embodiment, a cervical secretion sample includes, for example, a sample comprising any cervical secretion or secretions from areas anatomically within the vicinity of the cervix (e.g., vaginal secretion, endometrial secretion, and ovarian secretion, etc.). In some embodiments, the collection device has an absorbing member capable of collecting a sample (e.g., tissue, secretions, and/or cells) upon contact with a bodily region (e.g., cervix, vaginal canal). In some embodiments, the absorbing member is a sponge having a shape and size suitable for insertion into a body orifice (e.g., cervix, vaginal canal), for example having a cylindrical shape In some embodiments, the collection device is a tampon (e.g., a standard tampon), a lavage that releases liquid into the vagina and re-collects fluid (e.g., a Pantarhei screener), a cervical brush (e.g., a brush inserted into the vagina and turned around to collect cells), a Fournier cervical self-sampling device (a tampon-like plastic wand), or a swab. In some embodiments, the absorbing member is made of a material capable of collecting the desired sample. In some embodiments, the absorbing member is a sponge material, such as rayon and/or cotton.

Cell free or substantially cell free samples can be obtained by subjecting the sample to various techniques known to those of skill in the art which include, but are not limited to, centrifugation and filtration. Although it is generally preferred that no invasive techniques are

used to obtain the sample, it still may be preferable to obtain samples such as tissue homogenates, tissue sections, and biopsy specimens. The technology is not limited in the methods used to prepare the samples and provide a nucleic acid for testing. For example, in some embodiments, a DNA is isolated from a sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) using direct gene capture, *e.g.*, as detailed in U.S. Pat. Nos. 8,808,990 and 9,169,511, and in WO 2012/155072, or by a related method.

The analysis of markers can be carried out separately or simultaneously with additional markers within one test sample. For example, several markers can be combined into one test for efficient processing of multiple samples and for potentially providing greater diagnostic and/or prognostic accuracy. In addition, one skilled in the art would recognize the value of testing multiple samples (for example, at successive time points) from the same subject. Such testing of serial samples can allow the identification of changes in marker methylation states over time. Changes in methylation state, as well as the absence of change in methylation state, can provide useful information about the disease status that includes, but is not limited to, identifying the approximate time from onset of the event, the presence and amount of salvageable tissue, the appropriateness of drug therapies, the effectiveness of various therapies, and identification of the subject's outcome, including risk of future events. The analysis of biomarkers can be carried out in a variety of physical formats. For example, the use of microtiter plates or automation can be used to facilitate the processing of large numbers of test samples. Alternatively, single sample formats could be developed to facilitate immediate treatment and diagnosis in a timely fashion, for example, in ambulatory transport or emergency room settings.

Genomic DNA may be isolated by any means, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants, e.g., by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction, or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense, and required quantity of DNA. All clinical sample types comprising neoplastic matter or pre-neoplastic matter are suitable for use in the present

method, e.g., tissue (e.g., cervical tissue), cell lines, histological slides, biopsies, paraffin-embedded tissue, secretions (e.g., cervical secretions, vaginal secretions), body fluids, stool, tissue, colonic effluent, urine, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood, and combinations thereof.

The technology is not limited in the methods used to prepare the samples and provide a nucleic acid for testing. For example, in some embodiments, a DNA is isolated from a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) using direct gene capture, e.g., as detailed in U.S. Pat. Appl. Ser. No. 61/485386 or by a related method.

The genomic DNA sample is then treated with at least one reagent, or series of reagents, that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker comprising a DMR (e.g., DMR 1–423, e.g., as provided by Tables I, III, and X).

In some embodiments, the reagent converts cytosine bases which are unmethylated at the 5′-position to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. However in some embodiments, the reagent may be a methylation sensitive restriction enzyme.

In some embodiments, the genomic DNA sample is treated in such a manner that cytosine bases that are unmethylated at the 5′ position are converted to uracil, thymine, or another base that is dissimilar to cytosine in terms of hybridization behavior. In some embodiments, this treatment is carried out with bisulfite (hydrogen sulfite, disulfite) followed by alkaline hydrolysis.

The treated nucleic acid is then analyzed to determine the methylation state of the target gene sequences (at least one gene, genomic sequence, or nucleotide from a marker comprising a DMR, e.g., at least one DMR chosen from DMR 1–423, e.g., as provided in Tables I, III, and X). The method of analysis may be selected from those known in the art, including those listed herein, e.g., QuARTS and MSP as described herein.

Aberrant methylation, more specifically hypermethylation of a marker comprising a DMR (e.g., DMR 1–423, e.g., as provided by Tables I, III, and X) is associated with cervical cancer, cervical cancer subtypes, and cervical pre-cancers.

In some embodiments, the technology relates to a method for treating a patient (e.g., a patient with any cervical cancer and/or a cervical cancer subtype), the method comprising determining either or both of 1) the methylation state of one or more methylation marker as provided herein, and administering a treatment to the patient based on the results of determining the methylation state. The treatment may be administration of a pharmaceutical compound, a vaccine, performing a surgery, imaging the patient, performing another test. Preferably, said use is in a method of clinical screening, a method of prognosis assessment, a method of monitoring the results of therapy, a method to identify patients most likely to respond to a particular therapeutic treatment, a method of imaging a patient or subject, and a method for drug screening and development.

In some embodiments of the technology, a method for diagnosing a specific type of cancer (e.g., cervical cancer or a subtype of cervical cancer) and/or pre-cancer (e.g., cervical pre-cancer) in a subject is provided. The terms "diagnosing" and "diagnosis" as used herein refer to methods by which the skilled artisan can estimate and even determine whether or not a subject is suffering from a given disease or condition or may develop a given disease or condition in the future. The skilled artisan often makes a diagnosis on the basis of one or more diagnostic indicators, such as for example one or more biomarkers (e.g., one or more methylated markers, methylated marker genes, genes, DMRs, and/or DNA methylated markers as disclosed herein), the methylation state of which is indicative of the presence, severity, or absence of the condition.

Along with diagnosis, clinical cancer prognosis relates to determining the aggressiveness of the cancer and the likelihood of tumor recurrence to plan the most effective therapy. If a more accurate prognosis can be made or even a potential risk for developing the cancer can be assessed, appropriate therapy, and in some instances less severe therapy for the patient can be chosen. Assessment (e.g., determining methylation state) of cancer biomarkers is useful to separate subjects with good prognosis and/or low risk of developing cancer (e.g., cervical cancer or a subtype of cervical cancer) who will need no therapy or limited therapy from those more likely to develop cancer or suffer a recurrence of cancer who might benefit from more intensive treatments.

As such, "making a diagnosis" or "diagnosing", as used herein, is further inclusive of determining a risk of developing cancer (e.g., cervical cancer or a subtype of cervical cancer) or determining a prognosis, which can provide for predicting a clinical outcome (with or without medical treatment), selecting an appropriate treatment (or whether treatment would

be effective), or monitoring a current treatment and potentially changing the treatment, based on the measure of the diagnostic biomarkers (e.g., DMR) disclosed herein. Further, in some embodiments of the presently disclosed subject matter, multiple determination of the biomarkers over time can be made to facilitate diagnosis and/or prognosis. A temporal change in the biomarker can be used to predict a clinical outcome, monitor the progression of cancer or a subtype of cancer (e.g., cervical cancer or a subtype of cervical cancer), and/or monitor the efficacy of appropriate therapies directed against the cancer. In such an embodiment for example, one might expect to see a change in the methylation state of one or more biomarkers (e.g., DMR) disclosed herein (and potentially one or more additional biomarker(s), if monitored) in a biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) over time during the course of an effective therapy.

The presently disclosed subject matter further provides in some embodiments a method for determining whether to initiate or continue prophylaxis or treatment of a cancer (e.g., cervical cancer or a subtype of cervical cancer) in a subject. Any changes over the time period can be used to predict risk of developing cancer (e.g., cervical cancer or a subtype of cervical cancer), predict clinical outcome, determine whether to initiate or continue the prophylaxis or therapy of the cancer, and whether a current therapy is effectively treating the cancer. For example, a first time point can be selected prior to initiation of a treatment and a second time point can be selected at some time after initiation of the treatment. Methylation states can be measured in each of the samples taken from different time points and qualitative and/or quantitative differences noted. A change in the methylation states of the biomarker levels from the different samples can be correlated with a specific cancer (e.g., cervical cancer or a subtype of cervical cancer) risk, prognosis, determining treatment efficacy, and/or progression of the cancer in the subject.

In preferred embodiments, the methods and compositions of the invention are for treatment or diagnosis of disease (e.g., cervical cancer or a subtype of cervical cancer) at an early stage, for example, before symptoms of the disease appear. In some embodiments, the methods and compositions of the invention are for treatment or diagnosis of disease at a clinical stage.

As noted, in some embodiments, multiple determinations of one or more diagnostic or prognostic biomarkers can be made, and a temporal change in the marker can be used to

determine a diagnosis or prognosis. For example, a diagnostic marker can be determined at an initial time, and again at a second time. In such embodiments, an increase in the marker from the initial time to the second time can be diagnostic of a particular type or severity of cancer (e.g., cervical cancer or a subtype of cervical cancer), or a given prognosis. Likewise, a decrease in the marker from the initial time to the second time can be indicative of a particular type or severity of cancer, or a given prognosis. Furthermore, the degree of change of one or more markers can be related to the severity of the cancer and future adverse events. The skilled artisan will understand that, while in certain embodiments comparative measurements can be made of the same biomarker at multiple time points, one can also measure a given biomarker at one time point, and a second biomarker at a second time point, and a comparison of these markers can provide diagnostic information.

As used herein, the phrase "determining the prognosis" refers to methods by which the skilled artisan can predict the course or outcome of a condition in a subject. The term "prognosis" does not refer to the ability to predict the course or outcome of a condition with 100% accuracy, or even that a given course or outcome is predictably more or less likely to occur based on the methylation state of a biomarker (e.g., a DMR). Instead, the skilled artisan will understand that the term "prognosis" refers to an increased probability that a certain course or outcome will occur; that is, that a course or outcome is more likely to occur in a subject exhibiting a given condition, when compared to those individuals not exhibiting the condition. For example, in individuals not exhibiting the condition (e.g., having a normal methylation state of one or more DMR), the chance of a given outcome (e.g., suffering from a specific type of cancer (e.g., cervical cancer or a subtype of cervical cancer)) may be very low.

In some embodiments, a statistical analysis associates a prognostic indicator with a predisposition to an adverse outcome. For example, in some embodiments, a methylation state different from that in a normal control sample obtained from a patient who does not have a cancer (e.g., cervical cancer or a subtype of cervical cancer) can signal that a subject is more likely to suffer from a cancer than subjects with a level that is more similar to the methylation state in the control sample, as determined by a level of statistical significance. Additionally, a change in methylation state from a baseline (e.g., "normal") level can be reflective of subject prognosis, and the degree of change in methylation state can be related to the severity of adverse events. Statistical significance is often determined by comparing two or more populations and determining a confidence interval and/or a *p* value. See, e.g., Dowdy

and Wearden, Statistics for Research, John Wiley & Sons, New York, 1983, incorporated herein by reference in its entirety. Exemplary confidence intervals of the present subject matter are 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% and 99.99%, while exemplary *p* values are 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, and 0.0001.

In other embodiments, a threshold degree of change in the methylation state of a prognostic or diagnostic biomarker disclosed herein (e.g., a DMR) can be established, and the degree of change in the methylation state of the biomarker in a biological sample is simply compared to the threshold degree of change in the methylation state. A preferred threshold change in the methylation state for biomarkers provided herein is about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 50%, about 75%, about 100%, and about 150%. In yet other embodiments, a "nomogram" can be established, by which a methylation state of a prognostic or diagnostic indicator (biomarker or combination of biomarkers) is directly related to an associated disposition towards a given outcome. The skilled artisan is acquainted with the use of such nomograms to relate two numeric values with the understanding that the uncertainty in this measurement is the same as the uncertainty in the marker concentration because individual sample measurements are referenced, not population averages.

In some embodiments, a control sample is analyzed concurrently with the biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample), such that the results obtained from the biological sample can be compared to the results obtained from the control sample. Additionally, it is contemplated that standard curves can be provided, with which assay results for the biological sample may be compared. Such standard curves present methylation states of a biomarker as a function of assay units, e.g., fluorescent signal intensity, if a fluorescent label is used. Using samples taken from multiple donors, standard curves can be provided for control methylation states of the one or more biomarkers in normal tissue, as well as for "at-risk" levels of the one or more biomarkers in plasma taken from donors with a specific type of cancer (e.g., cervical cancer or a subtype of cervical cancer) or pre-cancer (e.g., cervical pre-cancer). In certain embodiments of the method, a subject is identified as having cancer (e.g., cervical cancer or a subtype of cervical cancer) or a pre-cancer (e.g., cervical pre-cancer) upon identifying an aberrant methylation state of one or more DMR provided herein in a biological sample obtained from the subject. In other

embodiments of the method, the detection of an aberrant methylation state of one or more of such biomarkers in a biological sample obtained from the subject results in the subject being identified as having cancer (e.g., cervical cancer or a subtype of cervical cancer) or a pre-cancer (e.g., cervical pre-cancer).

The analysis of markers can be carried out separately or simultaneously with additional markers within one test sample. For example, several markers can be combined into one test for efficient processing of a multiple of samples and for potentially providing greater diagnostic and/or prognostic accuracy. In addition, one skilled in the art would recognize the value of testing multiple samples (for example, at successive time points) from the same subject. Such testing of serial samples can allow the identification of changes in marker methylation states over time. Changes in methylation state, as well as the absence of change in methylation state, can provide useful information about the disease status that includes, but is not limited to, identifying the approximate time from onset of the event, the presence and amount of salvageable tissue, the appropriateness of drug therapies, the effectiveness of various therapies, and identification of the subject's outcome, including risk of future events.

The analysis of biomarkers can be carried out in a variety of physical formats. For example, the use of microtiter plates or automation can be used to facilitate the processing of large numbers of test samples. Alternatively, single sample formats could be developed to facilitate immediate treatment and diagnosis in a timely fashion, for example, in ambulatory transport or emergency room settings.

In some embodiments, the subject is diagnosed as having cervical cancer, a cervical cancer subtype, and/or a cervical pre-cancer if, when compared to a control methylation state, there is a measurable difference in the methylation state of at least one biomarker in the sample. Conversely, when no change in methylation state is identified in the biological sample, the subject can be identified as not having a cervical cancer, a cervical cancer subtype, and/or a cervical pre-cancer, not being at risk for the cancer or pre-cancer, or as having a low risk of the cancer or pre-cancer. In this regard, subjects having the cancer or risk thereof can be differentiated from subjects having low to substantially no cancer or risk thereof. Those subjects having a risk of developing a cervical cancer, a cervical cancer subtype, and/or a cervical pre-cancer can be placed on a more intensive and/or regular screening schedule. On the other hand, those subjects having low to substantially no risk may avoid being subjected to additional testing for cancer risk (e.g., invasive procedure), until

such time as a future screening, for example, a screening conducted in accordance with the present technology, indicates that a risk of cancer risk has appeared in those subjects.

As mentioned above, depending on the embodiment of the method of the present technology, detecting a change in methylation state of the one or more biomarkers can be a qualitative determination or it can be a quantitative determination. As such, the step of diagnosing a subject as having, or at risk of developing, cervical cancer or a cervical cancer subtype indicates that certain threshold measurements are made, e.g., the methylation state of the one or more biomarkers in the biological sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) varies from a predetermined control methylation state. In some embodiments of the method, the control methylation state is any detectable methylation state of the biomarker. In other embodiments of the method where a control sample is tested concurrently with the biological sample, the predetermined methylation state is the methylation state in the control sample. In other embodiments of the method, the predetermined methylation state is based upon and/or identified by a standard curve. In other embodiments of the method, the predetermined methylation is a specifically state or range of state. As such, the predetermined methylation state can be chosen, within acceptable limits that will be apparent to those skilled in the art, based in part on the embodiment of the method being practiced and the desired specificity, etc.

In certain embodiments, the technology provides steps for reacting a nucleic acid comprising a DMR with a reagent capable of modifying nucleic acid in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, Ten Eleven Translocation (TET) enzyme (e.g., human TET1, human TET2, human TET3, murine TET1, murine TET2, murine TET3, Naegleria TET (NgTET), Coprinopsis cinerea (CcTET)), or a variant thereof), borane reducing agent) to produce, for example, nucleic acid modified in a methylation-specific manner; sequencing the nucleic acid modified in a methylation-specific manner to provide a nucleotide sequence of the nucleic acid modified in a methylation-specific manner; comparing the nucleotide sequence of the nucleic acid modified in a methylation-specific manner with a nucleotide sequence of a nucleic acid comprising the DMR from a subject who does not have cervical cancer or a cervical cancer subtype to identify differences in the two

sequences; and identifying the subject as having cervical cancer or a cervical cancer subtype when a difference is present.

The technology further provides compositions. In certain embodiments, the technology provides composition comprising a nucleic acid comprising a DMR and a bisulfite reagent. In certain embodiments, composition comprising a nucleic acid comprising a DMR and one or more oligonucleotide according to SEQ ID NOS 1-76 are provided. In certain embodiments, compositions comprising a nucleic acid comprising a DMR and a methylation-sensitive restriction enzyme are provided. In certain embodiments, compositions comprising a nucleic acid comprising a DMR and a polymerase are provided.

The technology further provides kits. The kits comprise embodiments of the compositions, devices, apparatuses, *etc.* described herein, and instructions for use of the kit. Such instructions describe appropriate methods for preparing an analyte from a sample, *e.g.*, for collecting a sample (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample) and preparing a nucleic acid from the sample. In some embodiments, the kits contain one or more collection devices capable of obtaining a sample (e.g. tissue, secretions, and/or cells) from or near the cervix (e.g., a tampon (e.g., a standard tampon), a lavage that releases liquid into the vagina and re-collects fluid (e.g., a Pantarhei screener), a cervical brush (e.g., a brush inserted into the vagina and turned around to collect cells), a Fournier cervical self-sampling device (a tampon-like plastic wand), or a swab). Individual components of the kit are packaged in appropriate containers and packaging (*e.g.*, vials, boxes, blister packs, ampules, jars, bottles, tubes, and the like) and the components are packaged together in an appropriate container (*e.g.*, a box or boxes) for convenient storage, shipping, and/or use by the user of the kit. It is understood that liquid components (*e.g.*, a buffer) may be provided in a lyophilized form to be reconstituted by the user. Kits may include a control or reference for assessing, validating, and/or assuring the performance of the kit. For example, a kit for assaying the amount of a nucleic acid present in a sample may include a control comprising a known concentration of the same or another nucleic acid for comparison and, in some embodiments, a detection reagent (*e.g.*, a primer) specific for the control nucleic acid. The kits are appropriate for use in a clinical setting and, in some embodiments, for use in a user's home. The components of a kit, in some embodiments,

provide the functionalities of a system for preparing a nucleic acid solution from a sample. In some embodiments, certain components of the system are provided by the user.

In certain embodiments, the technology is related to embodiments of compositions (e.g., reaction mixtures). In some embodiments are provided a composition comprising a nucleic acid comprising a DMR and a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, Ten Eleven Translocation (TET) enzyme (e.g., human TET1, human TET2, human TET3, murine TET1, murine TET2, murine TET3, Naegleria TET (NgTET), Coprinopsis cinerea (CcTET)), or a variant thereof), borane reducing agent). Some embodiments provide a composition comprising a nucleic acid comprising a DMR and an oligonucleotide as described herein. Some embodiments provide a composition comprising a nucleic acid comprising a DMR and a methylation-sensitive restriction enzyme. Some embodiments provide a composition comprising a nucleic acid comprising a DMR and a polymerase.

In some embodiments, the technology described herein is associated with a programmable machine designed to perform a sequence of arithmetic or logical operations as provided by the methods described herein. For example, some embodiments of the technology are associated with (e.g., implemented in) computer software and/or computer hardware. In one aspect, the technology relates to a computer comprising a form of memory, an element for performing arithmetic and logical operations, and a processing element (e.g., a microprocessor) for executing a series of instructions (e.g., a method as provided herein) to read, manipulate, and store data. In some embodiments, a microprocessor is part of a system for determining a methylation state (e.g., of one or more DMR, e.g., DMR 1-423 as provided in Tables I, III, and/or X); comparing methylation states; generating standard curves; determining a Ct value; calculating a fraction, frequency, or percentage of methylation; identifying a CpG island; determining a specificity and/or sensitivity of an assay or marker; calculating an ROC curve and an associated AUC; sequence analysis; all as described herein or is known in the art.. In some embodiments, a microprocessor is part of a system for determining a methylation state (e.g., of one or more DMR, e.g., DMR 1-423 as provided in Tables I, III, and/or X); comparing methylation states; generating standard curves; determining a Ct value; calculating a fraction, frequency, or percentage of methylation; identifying a CpG island; determining a specificity and/or sensitivity of an assay or marker;

calculating an ROC curve and an associated AUC; sequence analysis; all as described herein or is known in the art.

In some embodiments, a software or hardware component receives the results of multiple assays and determines a single value result to report to a user that indicates a cancer risk (e.g., cervical cancer or a subtype of cervical cancer) or pre-cancer risk (e.g., cervical pre-cancer) based on the results of the multiple assays (e.g., determining the methylation state of multiple DMR, e.g., as provided in Tables I, III, and X). Related embodiments calculate a risk factor based on a mathematical combination (e.g., a weighted combination, a linear combination) of the results from the multiple assays (e.g., determining the methylation state of multiple DMR, e.g., as provided in Tables I, III, and X). In some embodiments, the methylation state of a DMR defines a dimension and may have values in a multidimensional space and the coordinate defined by the methylation states of multiple DMR is a result, e.g., to report to a user, e.g., related to a cancer risk.

In some embodiments, the technology provided herein is associated with a plurality of programmable devices that operate in concert to perform a method as described herein. For example, in some embodiments, a plurality of computers (e.g., connected by a network) may work in parallel to collect and process data, e.g., in an implementation of cluster computing or grid computing or some other distributed computer architecture that relies on complete computers (with onboard CPUs, storage, power supplies, network interfaces, etc.) connected to a network (private, public, or the internet) by a conventional network interface, such as Ethernet, fiber optic, or by a wireless network technology.

For example, some embodiments provide a computer that includes a computer-readable medium. The embodiment includes a random access memory (RAM) coupled to a processor. The processor executes computer-executable program instructions stored in memory. Such processors may include a microprocessor, an ASIC, a state machine, or other processor, and can be any of a number of computer processors, such as processors from Intel Corporation of Santa Clara, California and Motorola Corporation of Schaumburg, Illinois. Such processors include, or may be in communication with, media, for example computer-readable media, which stores instructions that, when executed by the processor, cause the processor to perform the steps described herein.

Computers are connected in some embodiments to a network. Computers may also include a number of external or internal devices such as a mouse, a CD-ROM, DVD, a keyboard, a display, or other input or output devices. Examples of computers are personal

computers, digital assistants, personal digital assistants, cellular phones, mobile phones, smart phones, pagers, digital tablets, laptop computers, internet appliances, and other processor-based devices. In general, the computers related to aspects of the technology provided herein may be any type of processor-based platform that operates on any operating system, such as Microsoft Windows, Linux, UNIX, Mac OS X, etc., capable of supporting one or more programs comprising the technology provided herein. Some embodiments comprise a personal computer executing other application programs (e.g., applications). The applications can be contained in memory and can include, for example, a word processing application, a spreadsheet application, an email application, an instant messenger application, a presentation application, an Internet browser application, a calendar/organizer application, and any other application capable of being executed by a client device.

All such components, computers, and systems described herein as associated with the technology may be logical or virtual.

In certain embodiments, the technology provides systems for screening cervical cancer, a cervical cancer subtype, and/or a cervical pre-cancer in a sample obtained from a subject are provided by the technology. Exemplary embodiments of systems include, e.g., a system for screening for cervical cancer or a cervical cancer subtype in a sample obtained from a subject (e.g., a tissue sample (e.g., cervical tissue), a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample (e.g., cervical secretion, vaginal secretion), an organ secretion sample, a CSF sample, a saliva sample, a urine sample, or a stool sample), the system comprising:

an analysis component configured to one or both of determining the methylation state of one or more methylated markers in a sample,

a software component configured to compare the methylation state of the one or more methylated markers in the sample with a control sample or a reference sample recorded in a database, and

an alert component configured to alert a user of a cancer associated state.

In some embodiments, an alert is determined by a software component that receives the results from multiple assays (e.g., determining the methylation states of the one or more methylated markers) and calculating a value or result to report based on the multiple results.

Some embodiments provide a database of weighted parameters associated with each methylated marker provided herein for use in calculating a value or result and/or an alert to report to a user (e.g., such as a physician, nurse, clinician, etc.). In some embodiments all

results from multiple assays are reported. In some embodiments, one or more results are used to provide a score, value, or result based on a composite of one or more results from multiple assays that is indicative of a cancer risk in a subject. Such methods are not limited to particular methylation markers.

In such methods and systems, the one or more methylation markers comprise a base in a DMR selected from a group consisting of DMR 1–423 as provided in Tables I, III, and X.

In this detailed description of the various embodiments, for purposes of explanation, numerous specific details are set forth to provide a thorough understanding of the embodiments disclosed. One skilled in the art will appreciate, however, that these various embodiments may be practiced with or without these specific details. In other instances, structures and devices are shown in block diagram form. Furthermore, one skilled in the art can readily appreciate that the specific sequences in which methods are presented and performed are illustrative and it is contemplated that the sequences can be varied and still remain within the spirit and scope of the various embodiments disclosed herein.

EXAMPLES

Example I.

This example describes experiments conducted to assess the feasibility of targeted assay of a panel of methylated DNA markers (MDMs) for detection cervical cancer.

A proprietary methodology of sample preparation, sequencing, analyses pipelines, and filters were utilized to identify and narrow differentially methylated regions (DMRs) to those which would pinpoint cervical cancers and excel in a clinical testing environment. From the tissue to tissue analysis 320 hypermethylated CC DMRs were identified (Table I). Table II shows the area under the curve (AUC), fold-change, and p-value for cervical cancer tissue versus benign cervical tissue for the markers recited in Table I. The identified 320 hypermethylated CC DMRs included CC specific regions and CC subtype specific regions.

Table I. Identified methylated regions distinguishing cervical cancer tissue from benign cervical tissue (the genomic coordinates for the regions shown are based on the Human Feb. 2009 (GRCh37/hg19) Assembly)

DMR	Gene Annotation	Chromosome No.	Chromosome Coordinates
1	A1BG	19	58858561-58858983

2	ABCB1	7	87257074-87257239
3	ABCG5	2	44059004-44059094
4	ACCN1	17	31619175-31619259
5	ADAMTS3	4	73433788-73434198
6	AFF3	2	100721515-100721845
7	AK5	1	77747491-77747536
8	ALX4	11	44331425-44331467
9	ARHGAP12	10	32218791-32218902
10	ARHGAP20	11	110583038-110583345
11	ARL5C_A	17	37321102-37321336
12	ASCL1	12	103351856-103352059
13	ATP10A	15	26108318-26108663
14	AVPR1A	12	63545292-63545422
15	B4GALNT1_A	12	58026043-58026232
16	B4GALNT1_B	12	58026291-58026477
17	BARHL1	9	135462629-135462711
18	BCAT1	12	25101413-25101483
19	BEGAIN	14	101034327-101035263
20	BHLHE23	20	61638294-61638471
21	BMPER	7	33943944-33944081
22	BNIP3_A	10	133795098-133795395
23	BNIP3_B	10	133795980-133796111
24	C1orf114	1	169396696-169396795
25	C1orf200	1	9712256-9712761
26	C1orf95	1	226736897-226737246
27	C2orf40	2	106682149-106682286
28	CACNA1C	12	2162008-2162843
29	CACNA1G	17	48636452-48636851
30	CACNA2D3	3	54156377-54156417
31	CACNG8	19	54485957-54486134
32	CBFA2T3	16	89007193-89007582
33	CCDC39	3	180397115-180397545
34	CCDC63	12	111284678-111284903
35	CCND2_A	12	4378126-4378403
36	CCND2_B	12	4380227-4380491
37	CD1D	1	158150624-158151087
38	CD200	3	112052145-112052405
39	CD70	19	6590492-6590700
40	CDO1	5	115152309-115152413
41	CELF2	10	11207361-11207907
42	CHAT	10	50822420-50822511
43	CHST2	3	142839245-142839555
44	CLSTN2	3	139653864-139654264
45	CNRIP1	2	68546511-68546679
46	COLEC12	18	500692-500954
47	CPEB1_A	15	83316356-83316578
48	CPEB1_B	15	83316610-83316711
49	CPEB1_C	15	83316761-83316889
50	CRHR2	7	30721989-30722129
51	CTNND2	5	11903651-11903727
52	CYTH2	19	48984028-48984223
53	DFNA5	7	24796494-24796719

54	DPF3	14	73358910-73359002
55	DPP4	2	162930241-162930531
56	DPY19L2	12	64062412-64063072
57	DSCR6	21	38378602-38378731
58	DTX1	12	113494423-113494759
59	DYSF	2	71693487-71693650
60	ECEL1	2	233352371-233352699
61	EFNA2	19	1295824-1295929
62	ELMO1	7	37488818-37488937
63	EMID2	7	101006343-101006606
64	EOMES	3	27763266-27763325
65	EVC	4	5710271-5710311
66	EVI5L	19	7927499-7927549
67	FAIM2	12	50297738-50297953
68	FAM110B	8	59058124-59058186
69	FAM150A	8	53478223-53478377
70	FAM155A_A	13	108520148-108520265
71	FAM181B	11	82444358-82444592
72	FAM89A	1	231176282-231176575
73	FBN1_A	15	48936889-48936984
74	FBN1_B	15	48937320-48937731
75	FBN1_C	15	48938100-48938405
76	FEV	2	219849133-219849257
77	FGF3_A	11	69632118-69632678
78	FGF3_B	11	69633425-69633515
79	FLI1	11	128563956-128564293
80	FLJ32063	2	200328573-200328645
81	FLT1_A	13	29068764-29068896
82	FLT1_B	13	29069073-29069455
83	FLT3_A	13	28674335-28674415
84	FLT3_B	13	28674451-28674770
85	FOXP2	7	113725336-113725431
86	GALR1_A	18	74962194-74962252
87	GALR1_B	18	74962418-74962484
88	GBGT1	9	136038966-136039283
89	GDF6	8	97172702-97172754
90	GDF7	2	20866000-20866539
91	GLIS1	1	54204427-54205225
92	GLIS2	16	4377932-4378133
93	GPC6	13	93880086-93880188
94	GPM6A	4	176923218-176923302
95	GPR88	1	101004724-101004953
96	GREM2	1	240775358-240775401
97	GRIK3	1	37500263-37500335
98	GSX1	13	28363637-28363972
99	GYPC_A	2	127413591-127413703
100	GYPC_B	2	127413823-127413901
101	HLF	17	53343530-53343591
102	HOPX_A	4	57521687-57521903
103	HOPX_B	4	57522083-57522309
104	HOPX_C	4	57522385-57522479
105	HOPX_D	4	57522507-57522621

106	HOXA11	7	27227694-27227736
107	HRH2	5	175085354-175085467
108	IGSF9B_A	11	133826275-133826414
109	IGSF9B_B	11	133826934-133827316
110	IKZF1	7	50343331-50343552
111	IQUB	7	123172894-123173074
112	IRF4_A	6	391781-391900
113	IRF4_B	6	392442-392483
114	IRF4_C	6	393540-393626
115	ITGA4	2	182321830-182322266
116	ITGA5	12	54812290-54812533
117	ITPKA	15	41793563-41793621
118	ITPKB	1	226924925-226925000
119	JAM2	21	27011766-27012116
120	JAM3_A	11	133938788-133938917
121	JAM3_B	11	133939006-133939134
122	JSRP1_A	19	2253228-2253345
123	KATNAL2	18	44526733-44526773
124	KCNA1	12	5019366-5019433
125	KCNA3_A	1	111217011-111217092
126	KCNA3_B	1	111217621-111217859
127	KCNK12_A	2	47796802-47796931
128	KCNK12_B	2	47797362-47797417
129	KCNK17	6	39281389-39281585
130	KCNK9_A	8	140715057-140715148
131	KCNK9_B	8	140716165-140716252
132	KCNQ5	6	73331613-73331787
133	KIAA1383	1	232941228-232941421
134	LBH	2	30453651-30454103
135	LOC100129620	1	99470477-99470580
136	LOC100132891	8	72756111-72756258
137	LOC100192379	4	122686140-122686432
138	LOC157627_A	8	9763927-9764066
139	LOC157627_B	8	9764330-9764450
140	LOC157627_C	8	9764461-9764505
141	LOC220930	10	31608483-31609348
142	LOC642345	13	88323596-88323783
143	LOC644189	19	36909224-36909659
144	LOC648809	15	84748985-84749166
145	LPHN1	19	14260386-14260651
146	LPPR3	19	821418-821789
147	LY6H	8	144241411-144241518
148	MATK	19	3785879-3786193
149	MAX.chr1.161591532-161591616	1	161591532-161591616
150	MAX.chr1.228652332-228652455	1	228652332-228652455
151	MAX.chr1.241587394-241587493	1	241587394-241587493
152	MAX.chr1.35394602-35395059	1	35394602-35395059
153	MAX.chr1.98510958-98511049	1	98510958-98511049
154	MAX.chr10.102497254-102497366	10	102497254-102497366
155	MAX.chr10.131769903-131770042	10	131769903-131770042
156	MAX.chr11.14926602-14926647	11	14926602-14926647
157	MAX.chr11.57250204-57250611	11	57250204-57250611

158	MAX.chr11.58903539-58903592	11	58903539-58903592
159	MAX.chr11.59323785-59323833	11	59323785-59323833
160	MAX.chr12.4273874-4274123	12	4273874-4274123
161	MAX.chr12.52652294-52652357	12	52652294-52652357
162	MAX.chr12.53108215-53108272	12	53108215-53108272
163	MAX.chr13.25116339-25116387	13	25116339-25116387
164	MAX.chr13.29106835-29106997	13	29106835-29106997
165	MAX.chr14.96342482-96342588	14	96342482-96342588
166	MAX.chr15.28351832-28352241	15	28351832-28352241
167	MAX.chr15.78112404-78112692	15	78112404-78112692
168	MAX.chr17.45867384-45867662	17	45867384-45867662
169	MAX.chr17.8230314-8230459	17	8230314-8230459
170	MAX.chr18.73167725-73167817	18	73167725-73167817
171	MAX.chr19.17501437-17501524	19	17501437-17501524
172	MAX.chr19.20959229-20959643	19	20959229-20959643
173	MAX.chr19.30718424-30718720	19	30718424-30718720
174	MAX.chr19.4580599-4580736	19	4580599-4580736
175	MAX.chr19.4584907-4585088	19	4584907-4585088
176	MAX.chr2.105488688-105488830	2	105488688-105488830
177	MAX.chr2.127783183-127783403	2	127783183-127783403
178	MAX.chr2.173099703-173099999	2	173099703-173099999
179	MAX.chr2.45162181-45162420	2	45162181-45162420
180	MAX.chr2.97193223-97193287	2	97193223-97193287
181	MAX.chr2.97193452-97193624	2	97193452-97193624
182	MAX.chr20.21491441-21491503	20	21491441-21491503
183	MAX.chr20.34893992-34894061	20	34893992-34894061
184	MAX.chr20.58146884-58146954	20	58146884-58146954
185	MAX.chr20.62733800-62733905	20	62733800-62733905
186	MAX.chr22.42679482-42679979	22	42679482-42679979
187	MAX.chr22.50064415-50064560	22	50064415-50064560
188	MAX.chr22.50118517-50118677	22	50118517-50118677
189	MAX.chr3.14852716-14852812	3	14852716-14852812
190	MAX.chr3.28616834-28616874	3	28616834-28616874
191	MAX.chr3.69591689-69591784	3	69591689-69591784
192	MAX.chr4.41884120-41884180	4	41884120-41884180
193	MAX.chr4.8859853-8859939	4	8859853-8859939
194	MAX.chr5.42952182-42952307	5	42952182-42952307
195	MAX.chr5.77148578-77148655	5	77148578-77148655
196	MAX.chr5.77268554-77268725	5	77268554-77268725
197	MAX.chr5.87437130-87437457	5	87437130-87437457
198	MAX.chr6.130686783-130687268	6	130686783-130687268
199	MAX.chr6.58147682-58147771	6	58147682-58147771
200	MAX.chr7.121956750-121956806	7	121956750-121956806
201	MAX.chr7.155259633-155259737	7	155259633-155259737
202	MAX.chr7.1704248-1704556	7	1704248-1704556
203	MAX.chr7.1706132-1706343	7	1706132-1706343
204	MAX.chr7.63767351-63767404	7	63767351-63767404
205	MAX.chr8.30769438-30769726	8	30769438-30769726
206	MAX.chr8.688331-688393	8	688331-688393
207	MAX.chr9.99983863-99983910	9	99983863-99983910
208	MIAT	22	27053248-27053559
209	MUC12	7	100609556-100609627

210	NALCN	13	102068469-102068567
211	NCAM1	11	112833899-112834016
212	NEGR1	1	72747781-72747847
213	NEUROG3	10	71332209-71332333
214	NID2	14	52535717-52536126
215	NPAS1	19	47523654-47523718
216	NT5C1A	1	40137279-40137914
217	NTRK3_A	15	88799876-88800380
218	NTRK3_B	15	88800388-88800700
219	NXPH1	7	8473513-8473626
220	OLIG1	21	34442361-34442433
221	PARVB	22	44420554-44420623
222	PAX2	10	102588392-102588514
223	PAX5	9	37034459-37034563
224	PCDH20	13	61987998-61988042
225	PCDH9	13	67804510-67804731
226	PDE4B	1	66798105-66798275
227	PDGFD	11	104034774-104034920
228	PDGFRA	4	55099048-55099343
229	PIF1	15	65116475-65116558
230	PLCL1	2	198669886-198670352
231	PLEKHO1	1	150122962-150123146
232	POMC	2	25391018-25391218
233	PRDM12	9	133536476-133536577
234	PRKCG	19	54410112-54410197
235	PTENP1	9	33676793-33676938
236	PTGDR_A	14	52734526-52734775
237	PTGDR_B	14	52735213-52735395
238	PTPRM	18	7568044-7568115
239	PTPRU	1	29586341-29586390
240	RFX4	12	106979881-106979929
241	RORB	9	77111792-77112062
242	RYR3	15	33603624-33603787
243	SALL3	18	76739229-76739404
244	Septin9_A	17	75370102-75370194
245	Septin9_B	17	75370525-75370663
246	SFMBT2_A	10	7450246-7450327
247	SFMBT2_B	10	7452029-7452478
248	SHANK3_A	22	51112191-51112399
249	SHANK3_B	22	51112422-51112586
250	SLC24A4	14	92790534-92790575
251	SLC26A10	12	58015554-58015696
252	SLC35F1	6	118228394-118228489
253	SLC6A3	5	1445473-1445666
254	SLC9A3	5	528714-528778
255	SNX32	11	65601128-65601301
256	SPOCK2	10	73847889-73848052
257	ST6GALNAC5	1	77334046-77334125
258	ST8SIA1	12	22486883-22487168
259	ST8SIA3	18	55021390-55021467
260	STX1B	16	31021723-31022164
261	SUSD5	3	33260131-33260258

262	SYNE1	6	152958097-152958463
263	SYT15	10	46970720-46970775
264	SYT6	1	114695532-114695720
265	TET1	10	70320226-70320873
266	TJP2	9	71789424-71789541
267	TLX2	2	74741203-74741284
268	TMC2	20	2539517-2539610
269	TMEFF2	2	193059999-193060210
270	TMEM178	2	39892822-39893199
271	TMEM200C	18	5890757-5890849
272	TRIM15	6	30139641-30139766
273	TRIM58	1	248020399-248020450
274	TRIM71	3	32859744-32859793
275	TRIM9	14	51561868-51562422
276	TRPC3_A	4	122872067-122872241
277	TRPC3_B	4	122872703-122873038
278	TSHZ3_A	19	31839734-31840137
279	TSHZ3_B	19	31841190-31841535
280	TSPAN11	12	31079362-31079640
281	TTYH1	19	54926696-54926845
282	ULBP1	6	150285465-150285546
283	UTF1	10	135043803-135043880
284	VILL_A	3	38035507-38035743
285	VILL_B	3	38035975-38036061
286	VSTM2B_A	19	30016244-30016358
287	VSTM2B_B	19	30017444-30017485
288	VSX1	20	25065255-25065331
289	WDR17	4	176987103-176987198
290	WNT3	17	44896147-44896284
291	XKR6	8	10872819-10873457
292	ZBTB16	11	113929791-113930273
293	ZFP41	8	144328582-144328648
294	ZIK1	19	58095616-58095660
295	ZMIZ1_A	10	81002818-81002953
296	ZMIZ1_B	10	81003082-81003162
297	ZNF132	19	58951402-58951453
298	ZNF134	19	58125542-58125779
299	ZNF304	19	57862463-57863095
300	ZNF382	19	37095959-37096132
301	ZNF419	19	57999105-57999506
302	ZNF43	19	22018452-22018947
303	ZNF470	19	57078657-57078833
304	ZNF530	19	58111265-58111624
305	ZNF549	19	58038983-58039279
306	ZNF568	19	37407214-37407365
307	ZNF583	19	56915358-56915921
308	ZNF586_A	19	58280987-58281064
309	ZNF586_B	19	58281140-58281369
310	ZNF69	19	11998671-11998972
311	ZNF701	19	53073536-53073713
312	ZNF737	19	20748263-20748472
313	ZNF763	19	12075727-12076043

314	ZNF773	19	58011327-58011598
316	ZNF776	19	58258169-58258561
317	ZNF781	19	38182950-38183230
318	ZNF844	19	12175522-12175663
319	ZNF85	19	21106043-21106387
320	ZSCAN18	19	58609761-58609888

Table II. Area under the curve (AUC), fold-change (FC), and p-value for cervical cancer tissue versus benign cervical tissue (Normal) for the markers recited in Table I.

DMR	Gene Annotation	AUC Cervical Cancer.vs.Normal	FC Cervical Cancer.vs.Normal	p-value Cervical Cancer.vs.Normal
1	A1BG	0.8681	163.9	0.0007041
2	ABCB1	0.9062	225.3	4.30E-05
3	ABCG5	0.8715	216.2	0.04206
4	ACCN1	0.905	100.1	0.005458
5	ADAMTS3	0.8513	150.1	0.0339
6	AFF3	0.9007	332.5	0.005802
7	AK5	0.875	121.8	0.03173
8	ALX4	0.8556	128.8	0.01156
9	ARHGAP12	0.9187	29.39	0.001022
10	ARHGAP20	0.8824	90.6	0.01308
11	ARL5C_A	0.9575	110.5	0.001267
12	ASCL1	0.9127	145.7	0.00005461
13	ATP10A	0.923	178.3	0.001411
14	AVPR1A	0.8873	24.75	0.001443
15	B4GALNT1_A	0.9073	79.99	9.17E-05
16	B4GALNT1_B	0.9173	66.23	0.02039
17	BARHL1	0.9982	87.43	2.42E-07
18	BCAT1	0.8859	51.82	0.01044
19	BEGAIN	0.9297	147	0.03445
20	BHLHE23	0.8929	149.8	2.77E-06
21	BMPER	0.8705	588	0.04749
22	BNIP3_A	0.8599	53.82	0.04754
23	BNIP3_B	0.891	83.44	0.005236
24	C1orf114	0.9464	191.9	0.002913
25	C1orf200	0.8858	237.3	0.01615
26	C1orf95	0.9308	189.9	0.02143
27	C2orf40	0.9065	58.72	0.002932
28	CACNA1C	0.9755	238.7	0.007272
29	CACNA1G	0.9377	54.43	0.003661
30	CACNA2D3	0.8616	288.6	0.03224
31	CACNG8	0.9273	62.21	0.04407
32	CBFA2T3	0.9014	183.9	0.008917
33	CCDC39	0.9048	106.5	0.04371
34	CCDC63	0.8798	66.55	0.007124
35	CCND2_A	0.9135	52.33	0.0007264
36	CCND2_B	0.8711	45.64	0.006238
37	CD1D	0.8919	92.98	0.0002146
38	CD200	0.8685	113.4	0.02155
39	CD70	0.9127	37.58	0.02947

40	CDO1	0.9608	248.3	0.0002878
41	CELF2	0.8905	49.93	6.38E-05
42	CHAT	0.8503	95.68	0.002803
43	CHST2	0.9022	280.8	0.04599
44	CLSTN2	0.8806	206.3	0.006455
45	CNRIP1	0.9287	180.2	0.004064
46	COLEC12	0.8922	79.93	0.0001851
47	CPEB1_A	0.8708	51.97	0.001492
48	CPEB1_B	0.8595	71.87	0.04946
49	CPEB1_C	0.8937	37.77	0.01035
50	CRHR2	0.9299	127.5	0.009806
51	CTNND2	0.9053	47.93	3.55E-05
52	CYTH2	0.8975	25.22	1.69E-05
53	DFNA5	0.8564	88.33	0.02238
54	DPF3	0.8505	153.8	0.03496
55	DPP4	0.9273	69.34	0.008132
56	DPY19L2	0.8933	94.9	0.01712
57	DSCR6	0.8859	83.2	2.49E-05
58	DTX1	0.8633	293.7	0.02987
59	DYSF	0.8508	211.1	0.00192
60	ECEL1	0.8927	115.9	0.006306
61	EFNA2	0.8824	24.26	0.00152
62	ELMO1	0.9031	85.56	0.01097
63	EMID2	0.9073	95.76	0.006303
64	EOMES	0.8972	301.1	0.03415
65	EVC	0.8857	153	0.03326
66	EVI5L	0.8529	26.61	2.49E-06
67	FAIM2	0.954	97.37	9.08E-07
68	FAM110B	0.8253	30.67	0.006014
69	FAM150A	0.858	84.76	0.0004955
70	FAM155A_A	0.8672	135.3	1.64E-05
71	FAM181B	0.8654	255.1	0.01558
72	FAM89A	0.9213	36.37	0.005412
73	FBN1_A	0.8634	95.51	2.54E-05
74	FBN1_B	0.9299	266.2	0.001173
75	FBN1_C	0.891	253.6	0.007426
76	FEV	0.8859	93.65	0.001174
77	FGF3_A	0.934	116.1	0.004175
78	FGF3_B	0.8653	174.8	0.02558
79	FLI1	0.918	80.1	0.004199
80	FLJ32063	0.8851	57.12	7.70E-05
81	FLT1_A	0.8672	73.17	0.0006431
82	FLT1_B	0.8601	257.8	0.002949
83	FLT3_A	0.9118	45.13	0.0002446
84	FLT3_B	0.9152	75.26	0.0003903
85	FOXP2	0.8512	39.81	0.0004998
86	GALR1_A	0.8599	155.1	0.01261
87	GALR1_B	0.8872	99.43	0.006009
88	GBGT1	0.8927	87.98	0.0002011
89	GDF6	0.8814	97.61	9.11E-07
90	GDF7	0.8529	177.4	0.003736
91	GLIS1	0.904	110.8	0.02295

92	GLIS2	0.8503	146.3	0.03235
93	GPC6	0.9007	154.1	1.03E-08
94	GPM6A	0.8627	89.7	9.47E-05
95	GPR88	0.8966	108	9.18E-05
96	GREM2	0.9047	117.7	0.02786
97	GRIK3	0.869	120.2	1.08E-06
98	GSX1	0.8681	34.82	0.008509
99	GYPC_A	0.9549	46.74	4.47E-08
100	GYPC_B	0.9101	213.8	0.0006732
101	HLF	0.8501	34.39	0.01946
102	HOPX_A	0.891	83.98	0.0009785
103	HOPX_B	0.9069	178.4	0.005854
104	HOPX_C	0.9108	138.4	0.006714
105	HOPX_D	0.8655	57.32	1.56E-05
106	HOXA11	0.875	41.78	0.00165
107	HRH2	0.8997	76.84	0.0001794
108	IGSF9B_A	0.917	82.34	0.007127
109	IGSF9B_B	0.8979	158	0.001897
110	IKZF1	0.92	78.83	0.03592
111	IQUB	0.893	72.54	0.03801
112	IRF4_A	0.9245	138.3	0.0003516
113	IRF4_B	0.9003	184.6	0.000418
114	IRF4_C	0.8807	92.83	0.02831
115	ITGA4	0.9013	222.2	0.01869
116	ITGA5	0.8704	35.52	2.58E-05
117	ITPKA	0.8585	27.23	0.0001397
118	ITPKB	0.9057	54.73	0.02796
119	JAM2	0.9268	162	0.0143
120	JAM3_A	0.8594	205.6	0.01462
121	JAM3_B	0.8733	129.6	0.004121
122	JSRP1_A	0.8737	69.72	0.0001102
123	KATNAL2	0.875	68.06	0.01101

124	KCNA1	0.9066	169.7	0.000422
125	KCNA3_A	0.872	155.8	0.0002929
126	KCNA3_B	0.8548	95.19	6.61E-05
127	KCNK12_A	0.9048	202.2	0.02022
128	KCNK12_B	0.8729	47.09	0.001777
129	KCNK17	0.9171	108	0.002709
130	KCNK9_A	0.8841	75.1	0.001264
131	KCNK9_B	0.8567	69.72	0.003352
132	KCNQ5	0.9031	101.9	0.003513
133	KIAA1383	0.8607	88.92	0.0007306
134	LBH	0.9343	97.12	0.0002605
135	LOC100129620	0.9062	290.7	0.02926
136	LOC100132891	0.8729	120.4	0.0007793
137	LOC100192379	0.8529	130.7	4.65E-05
138	LOC157627_A	0.8644	109.4	0.0001663
139	LOC157627_B	0.8676	81.49	0.005897
140	LOC157627_C	0.8656	218.6	0.0007851
141	LOC220930	0.924	120.9	0.04857
142	LOC642345	0.8946	111.8	0.001642
143	LOC644189	0.9125	106.2	0.01375
144	LOC648809	0.8966	478.2	0.04067
145	LPHN1	0.9446	109.9	0.03741
146	LPPR3	0.9412	106.1	0.01488
147	LY6H	0.8931	42.29	0.0003825
148	MATK	0.9134	279.5	0.003352
149	MAX.chr1.161591532-161591616	0.9248	146.6	0.02991
150	MAX.chr1.228652332-228652455	0.8895	34.63	0.001126

151	MAX.chr1.241587394-241587493	0.8676	39.12	0.005924
152	MAX.chr1.35394602-35395059	0.8824	104.6	0.002789
153	MAX.chr1.98510958-98511049	0.9472	74.07	2.52E-06
154	MAX.chr10.102497254-102497366	0.8752	27.42	0.001174
155	MAX.chr10.131769903-131770042	0.8913	109.3	2.56E-06
156	MAX.chr11.14926602-14926647	0.9747	222.3	0.01192
157	MAX.chr11.57250204-57250611	0.8962	30.1	6.59E-05
158	MAX.chr11.58903539-58903592	0.8859	488.4	0.04909
159	MAX.chr11.59323785-59323833	0.8581	32.36	0.001669
160	MAX.chr12.4273874-4274123	0.8686	61.99	0.00166
161	MAX.chr12.52652294-52652357	0.9148	120.2	0.0001635
162	MAX.chr12.53108215-53108272	0.8879	79.99	0.02578
163	MAX.chr13.25116339-25116387	0.852	20.66	0.0229
164	MAX.chr13.29106835-29106997	0.9255	360.4	0.00597
165	MAX.chr14.96342482-96342588	0.8844	58.34	0.02044
166	MAX.chr15.28351832-28352241	0.861	215.6	0.03226
167	MAX.chr15.78112404-78112692	0.9055	24.62	0.0008102
168	MAX.chr17.45867384-45867662	0.9013	143.9	0.003555
169	MAX.chr17.8230314-8230459	0.8958	39.54	0.003105
170	MAX.chr18.73167725-73167817	0.8613	46	0.01328

171	MAX.chr19.17501437-17501524	0.8824	23.07	0.002121
172	MAX.chr19.20959229-20959643	0.9066	248	0.02103
173	MAX.chr19.30718424-30718720	0.8806	50.54	0.01945
174	MAX.chr19.4580599-4580736	0.9228	123.2	0.0269
175	MAX.chr19.4584907-4585088	0.9733	213.8	0.004035
176	MAX.chr2.105488688-105488830	0.8685	26.75	0.0004967
177	MAX.chr2.127783183-127783403	0.8868	133.3	0.007291
178	MAX.chr2.173099703-173099999	0.8633	48.69	0.006586
179	MAX.chr2.45162181-45162420	0.866	52.28	0.0005311
180	MAX.chr2.97193223-97193287	0.8667	43.5	0.003816
181	MAX.chr2.97193452-97193624	0.8708	65.02	4.31E-05
182	MAX.chr20.21491441-21491503	0.9103	115.9	0.001712
183	MAX.chr20.34893992-34894061	0.8698	113.1	0.04005
184	MAX.chr20.58146884-58146954	0.8841	71.37	0.0002409
185	MAX.chr20.62733800-62733905	0.9011	89.44	0.02824
186	MAX.chr22.42679482-42679979	0.9083	43.53	0.0004212
187	MAX.chr22.50064415-50064560	0.888	51.26	0.001201
188	MAX.chr22.50118517-50118677	0.8753	49.43	0.03994
189	MAX.chr3.148527716-14852812	0.9066	58.55	0.01096
190	MAX.chr3.286166834-28616874	0.8985	78.99	0.04249
191	MAX.chr3.695911689-69591784	0.9256	85.32	0.03526
192	MAX.chr4.418844120-41884180	0.8877	59.23	0.004574

193	MAX.chr4.8859853-8859939	0.9246	58.29	5.48E-05
194	MAX.chr5.42952182-42952307	0.8824	42.08	3.48E-07
195	MAX.chr5.77148578-77148655	0.8695	38.05	0.0007677
196	MAX.chr5.77268554-77268725	0.8879	28.07	3.68E-05
197	MAX.chr5.87437130-87437457	0.8519	53.85	0.0004652
198	MAX.chr6.13086783-130687268	0.8824	121.3	0.007539
199	MAX.chr6.58147682-58147771	0.8845	331.6	9.05E-06
200	MAX.chr7.12195750-121956806	0.8754	34.26	2.83E-06
201	MAX.chr7.155259633-155259737	0.8628	48.14	0.003016
202	MAX.chr7.1704248-1704556	0.8592	127	0.004151
203	MAX.chr7.1706132-1706343	0.8824	108.9	0.0008021
204	MAX.chr7.63767351-63767404	0.8507	97.4	0.005447
205	MAX.chr8.30769438-30769726	0.8702	146.2	0.01308
206	MAX.chr8.688331-688393	0.8695	41.94	0.000996
207	MAX.chr9.99983863-99983910	0.9188	162.4	0.002767
208	MIAT	0.9118	45.21	0.005443
209	MUC12	0.8578	209.5	0.02915
210	NALCN	0.857	51.5	0.004903
211	NCAM1	0.9044	188.9	0.03646
212	NEGR1	0.8529	171.2	0.04721
213	NEUROG3	0.9165	158.2	0.003677
214	NID2	0.8449	98.22	0.0004152
215	NPAS1	0.8591	84.86	0.01148
216	NT5C1A	0.9671	124.2	0.006341
217	NTRK3_A	0.8945	77.83	1.95E-05
218	NTRK3_B	0.9002	66.48	4.92E-06
219	NXPH1	0.8351	81.58	0.000002667

220	OLIG1	0.8538	88.52	0.008082
221	PARVB	0.9444	418	0.02117
222	PAX2	0.8754	36.95	1.24E-05
223	PAX5	0.8585	66.76	0.02582
224	PCDH20	0.9173	161.7	0.001513
225	PCDH9	0.893	328.9	0.02864
226	PDE4B	0.862	27.99	0.0004054
227	PDGFD	0.8706	443.5	0.006209
228	PDGFRA	0.8973	137.5	0.001659
229	PIF1	0.8529	55.36	0.004562
230	PLCL1	0.8547	115.3	0.0007377
231	PLEKHO1	0.9055	39.62	0.0383
232	POMC	0.8746	38.48	0.001239
233	PRDM12	0.9146	68.76	0.002603
234	PRKCG	0.8655	43.59	0.02049
235	PTENP1	0.8763	455	0.01436
236	PTGDR_A	0.8613	67.78	3.93E-05
237	PTGDR_B	0.9231	45.63	3.22E-06
238	PTPRM	0.869	131.4	0.002421
239	PTPRU	0.8883	120.6	0.003687
240	RFX4	0.8815	318.9	0.008332
241	RORB	0.8971	141	0.004515
242	RYR3	0.8754	85.35	0.0003078
243	SALL3	0.8952	42.15	1.83E-07
244	Septin9_A	0.8609	127.4	0.02107
245	Septin9_B	0.9608	140.1	0.008477
246	SFMBT2_A	0.9474	68.95	0.005568
247	SFMBT2_B	0.8717	128.1	0.001783
248	SHANK3_A	0.8975	120.4	0.001461

249	SHANK3_B	0.8824	85.32	0.002408
250	SLC24A4	0.8539	107.5	0.002063
251	SLC26A10	0.8979	73.4	0.0265
252	SLC35F1	0.8529	322.5	0.01256
253	SLC6A3	0.8558	136.1	1.67E-06
254	SLC9A3	0.9479	26.3	0.003185
255	SNX32	0.9483	270.9	0.04352
256	SPOCK2	0.8851	128.9	0.02648
257	ST6GALNAC5	0.9195	169.3	2.43E-05
258	ST8SIA1	0.9585	143.1	1.96E-05
259	ST8SIA3	0.9273	88.4	0.0004173
260	STX1B	0.8616	55.36	3.55E-05
261	SUSD5	0.8768	90.17	0.01131
262	SYNE1	0.8788	94.57	0.01375
263	SYT15	0.9064	42.78	0.0494
264	SYT6	0.8868	136.1	0.01021
265	TET1	0.8997	179.4	0.02966
266	TJP2	0.8663	61.93	0.009941
267	TLX2	0.8979	70.83	0.01152
268	TMC2	0.9772	130.4	0.001039
269	TMEFF2	0.8755	58.54	1.57E-06
270	TMEM178	0.8841	47.63	0.001044
271	TMEM200C	0.9655	216.8	0.005405
272	TRIM15	0.943	185.6	0.03524
273	TRIM58	0.9725	95.12	0.0003784
274	TRIM71	0.8599	347.2	0.003817
275	TRIM9	0.8503	173.3	0.04287
276	TRPC3_A	0.8624	306.4	0.0004037
277	TRPC3_B	0.8642	389.8	0.04104

278	TSHZ3_A	0.9167	64.37	0.0007652
279	TSHZ3_B	0.8663	49.2	0.02169
280	TSPAN11	0.8877	211.8	0.03463
281	TTYH1	0.9421	122.1	0.02287
282	ULBP1	0.8859	66.25	1.53E-05
283	UTF1	0.9194	118.3	0.02302
284	VILL_A	0.9314	222.6	0.0003675
285	VILL_B	0.926	133.9	0.001691
286	VSTM2B_A	0.8704	106	0.000169
287	VSTM2B_B	0.9933	134	0.01423
288	VSX1	0.8786	58.31	3.46E-08
289	WDR17	0.8994	132.6	0.01159
290	WNT3	0.8636	114.5	0.02764
291	XKR6	0.8578	78.88	0.0005239
292	ZBTB16	0.9074	48.02	8.33E-06
293	ZFP41	0.9492	29.94	0.001962
294	ZIK1	0.8545	71.77	0.00014
295	ZMIZ1_A	0.8702	89.3	0.02504
296	ZMIZ1_B	0.9221	192.4	0.02884
297	ZNF132	0.8917	233.1	0.03695
298	ZNF134	0.8502	628.8	0.008461
299	ZNF304	0.917	231.5	0.004078
300	ZNF382	0.9152	126.8	0.003379
301	ZNF419	0.9002	301	0.04685
302	ZNF43	0.8849	161.7	0.01134
303	ZNF470	0.8734	578.3	0.02413
304	ZNF530	0.8962	113.4	0.04597
305	ZNF549	0.8636	92.32	0.0001819
306	ZNF568	0.8722	68.63	0.001613

307	ZNF583	0.8676	270.6	0.04555
308	ZNF586_A	0.9073	78.46	0.01952
309	ZNF586_B	0.8619	269.6	0.02721
310	ZNF69	0.9377	56.69	0.000363
311	ZNF701	0.8574	233.8	0.03808
312	ZNF737	0.9066	310	0.02542
313	ZNF763	0.8824	89.22	0.007205
314	ZNF773	0.9376	134.5	0.0003857
316	ZNF776	0.8676	43.03	0.01003
317	ZNF781	0.8841	386.9	0.001581
318	ZNF844	0.8832	122.8	0.04092
319	ZNF85	0.8503	120.1	0.01831
320	ZSCAN18	0.9239	81	0.04061

The tissue to leukocyte (buffy coat) analysis yielded 41 hypermethylated cervical tissue DMRs with less than 1% noise in WBCs (Table III, Table IV).

Table III. Identified methylated regions distinguishing cervical cancer tissue from leukocyte (buffy coat) (the genomic coordinates for the regions shown are based on the Human Feb. 2009 (GRCh37/hg19) Assembly)

DMR	Gene Annotation	Chromosome No.	Chromosome Coordinates
321	AGPAT3	21	45336891-45337191
322	AGRN	1	969276-969322
323	ARL5C_B	17	37321559-37321723
324	BCL2L11	2	111876440-111876822
325	BZRAP1	17	56406236-56406457
326	CA3	8	86350671-86350862
327	CALCA	11	14995338-14995473
328	CCDC88B	11	64108152-64108329
329	CD93	20	23066944-23067181
330	DLX5	7	96650568-96650655
331	EMX1	2	73147439-73147538
332	EMX1	2	73147887-73147940
333	FAM155A_B	13	108519903-108520040
334	GATA4	8	11565295-11565473
335	HLX	1	221050491-221050533

336	HOXD4	2	177017193-177017267
337	JSRP1_B	19	2252447-2252627
338	MAX.chr10.102900288-102900393	10	102900288-102900393
339	MAX.chr18.44780825-44780880	18	44780825-44780880
340	MAX.chr19.16394457-16394536	19	16394457-16394536
341	MAX.chr2.235355101-235355212	2	235355101-235355212
342	MAX.chr3.138658547-138658804	3	138658547-138658804
343	MAX.chr5.145725411-145725510	5	145725411-145725510
344	MAX.chr6.108440666-108440885	6	108440666-108440885
345	MAX.chr9.36739811-36739868	9	36739811-36739868
346	MAX.chr9.87904996-87905325	9	87904996-87905325
347	MIR196A1	17	46711093-46711156
348	MYF6	12	81102123-81102258
349	NCOR2	12	124950720-124950803
350	PKN1	19	14551093-14551303
351	PRIC285	20	62199258-62199703
352	PRMT7	16	68390231-68390556
353	psiTPTE22	22	17090731-17090824
354	SIX2	2	45232418-45232481
355	SORCS1	10	108924547-108924660
356	TBX5	12	114847371-114847489
357	TMEM132E	17	32964694-32964757
358	TNIK	3	171175939-171176064
359	TNK1	17	7287567-7287652
360	VSTM2B_C	19	30017026-30017110
361	ZC3H12D	6	149803483-149803682

Table IV. Area under the curve (AUC), fold-change (FC), and p-value for cervical cancer tissue versus leukocyte (buffy coat) for the markers recited in Table III.

DMR	Gene Annotation	AUC Cervical Cancer.vs. Buffy Coat	FC Cervical Cancer.vs. Buffy Coat	p-value Cervical Cancer.vs. Buffy Coat
321	AGPAT3	1	267.8	5.21E-06
322	AGRN	0.9596	195.9	0.000978
323	ARL5C_B	0.9753	125.5	0.000231
324	BCL2L11	0.9745	203.1	0.000338
325	BZRAP1	0.9552	161.1	0.000548
326	CA3	0.9562	264.9	0.000311
327	CALCA	0.9542	136.1	0.000885
328	CCDC88B	0.9696	125	0.000115
329	CD93	1	196	0.000108
330	DLX5	1	293.1	0.000771

331	EMX1	0.9562	252.4	0.000281
332	EMX1	0.9534	221.9	1.26E-08
333	FAM155A_B	0.9583	176.8	0.000519
334	GATA4	0.967	65.29	0.000473
335	HLX	1	252.8	0.000155
336	HOXD4	0.9615	124.3	1.12E-06
337	JSRP1_B	0.9526	78.12	3.43E-06
338	MAX.chr10.102900288-102900393	0.9512	62.76	0.000973
339	MAX.chr18.44780825-44780880	0.9606	62.88	0.000485
340	MAX.chr19.16394457-16394536	0.9545	290.9	0.000159
341	MAX.chr2.235355101-235355212	1	267.3	0.000389
342	MAX.chr3.138658547-138658804	0.9747	134	0.000255
343	MAX.chr5.145725411-145725510	0.9688	113.5	0.000119
344	MAX.chr6.108440666-108440885	0.9983	192.6	0.000562
345	MAX.chr9.3673981 1-36739868	0.9966	128	0.000668
346	MAX.chr9.87904996-87905325	0.9714	363	5.03E-09
347	MIR196A1	1	191.2	0.000149
348	MYF6	0.9869	50.03	1.38E-07
349	NCOR2	1	9719	3.27E-08
350	PKN1	0.9618	276.1	0.000263
351	PRIC285	0.9815	695	0.000469
352	PRMT7	1	370.5	1.57E-06
353	psiTPTE22	0.9545	97.69	0.000368
354	SIX2	1	497.3	0.000122
355	SORCS1	0.9755	175.1	0.000343
356	TBX5	0.9983	134.2	0.000536
357	TMEM132E	1	423	6.84E-07
358	TNIK	0.9531	303.1	7.77E-05
359	TNK1	0.9825	778.4	0.000116

360	VSTM2B_C	0.9545	134	0.000204
361	ZC3H12D	0.9512	91.08	8.80E-05

From these marker groups the following 29 candidates were chosen for an initial pilot: MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5. Quantitative methylation-specific PCR assays were developed and tested on independent samples. Short amplicon primers (<150bp) were designed to target the most discriminant CpGs with in a DMR and tested on controls to ensure that fully methylated fragments amplified robustly and in a linear fashion, that unmethylated and/or unconverted fragments did not amplify. The 60 primer sequences are listed in Table V.

Table V. Primer sequences for the 29 candidates chosen for initial analysis.

DMR#	Name	SEQ ID NO:	Forward Primer 5'-3' Sequence (hg19)	SEQ ID NO:	Reverse Primer 5'-3' Sequence (hg19)
2	ABCB1	1	CGCGAAAAGATTTTATATT GGTATTACGT	2	AACCCGCACCCGAAAC TAACGAC
9	ARHGAP12	3	TCG TAT TTT GTA AGG TAT AAT TCG G	4	AAC TCT TTC CTA TAC TTC ACT TCG TA
12	ASCL1	5	TAGAGTTTTTTATGCGTAG CGGCGG	6	ATTCTCTCTATATCCCC CTCGCGAA
13	ATP10A	7	TCG TGG TTA TCG CGG CGG AAA C	8	TAC CGC ACA AAA CCC CCT TAT CCT CG
13	ATP10A	9	TCG TGG TTA TCG CGG CGG AAA C	10	AAA CGC AAA CTA AAC GAA TAC CGC A
17	BARHL1	11	ATA TAA ACG TTA CGG GTA GGA GCG G	12	CAA CGA CAT ATC AAA ACC CGA CCT CG
24	C1orf114	13	TTT TAA TTA CGA GAG CGA TAA AAA TTT GCG T	14	CCG AAA TAA ATA CCG AAA AAA ATC GAT
27	C2orf40	15	CGG TTA GGG TTA GGA TAG TAG GTC GCG C	16	ATT CTC CCT CGC AAC ACC TCG AAA TAC G
28	CACNA1C	17	TTT CGA ATT TTG CGC GAA AGT CGT C	18	CGA CGA ATC TAA ATT AAT CCC TCC TCC GAA
50	CRHR2	19	GTT TTT GGG CGT TAT TTT CGG TCG T	20	TAC ACT CGA CGA CTC CTC TCC GAA
104	HOPX_C	21	GGA TAT AGT TTT TGT AAG GGG TTT CGG	22	TAT TAC GAA CTA ACG ACA AAA CCT AAC GTC
132	KCNQ5	23	AAG AGA AAT TTT TTT AAA GTG ACG T	24	AAA AAA AAC TAC AAA AAA AAC CGA A

153	MAX.chr1.98510968-98511049	25	ATT TTT TTT CGG AGA ATT CGA AAA GAA AAT ATG C	26	AAC TTC CAA ATC GAA AAT ATA ACG AA
170	MAX.chr18.73167751-73167791	27	TAG GAG GGG ACG TAG AGT TTA CGG CGA	28	GCG CAA CCC GAA CGA AAC GA
177	MAX.chr2.127783183-127783403	29	TTA GGT AGG ATT CGG ATG GCG AGG C	30	TCT ACA ACC GCA ACC AAT AAC GCC G
193	MAX.chr4.8859853-8859939	31	TTT TCG TTT ATT TGG GGG AAA TGG ATT TTC	32	ATC AAC ACA TCC GAA CCT CGC T
199	MAX.chr6.58147682-58147771	33	TGT TTA TGG ATT TAG GTG AGG ACG G	34	AAA AAA TAC AAC GTT TAA CCG CGA A
345	MAX.chr9.36739811-36739868	35	TGATAGGATGTTCGTTTA GTCGCGG	36	AAAAAACTACGCCGAT CCCCGAA
213	NEUROG3	37	GGT CGT TTT TTT AGC GAC GCG GC	38	CGA AAA ACT AAA CAA CCC AAA CGA T
214	NID2	39	AGG AGT TTT TAT TTC GTT AGT TTC GT	40	ATA ACC ACC ACA TCT AAT TCT CGT T
219	NXPH1	41	TAGTTCGCGAGAGTTTGA GAGTCGG	42	AACACGCCTACCTTCC TAACACGAA
233	PRDM12	43	AGG TTA TTT TCG TCG TTC GGG	44	ATA TCC CTC CGT AAA ACC GTC GAC C
254	SLC9A3	45	GCG TTT ACG GGG GAG GTC GT	46	TCT ACT ACG ATA TCT TAA CGA TCC GCG AAC
258	ST8SIA1	47	CGG TTG TTT AAC GAG AAA GAG ATC GT	48	GAT CTA ATT CCT CCT CCA CGC CGT A
271	TMEM200C	49	GAA CGC GGT TTT TAG GAG ATT TCG A	50	CAT ACT ACC CTC TAC GCC GCG AA
281	TTYH1	51	TTT TTG TTT TTG GGG CGC GAA AAC	52	ATA CAT CTC CTC CAC CAA CTA CCC CGC
300	ZNF382	53	GGG AGT CGG GGT TTT GGT AGA AGC G	54	ACA CCC ACG ACC GCC CTA TTA CGA C
310	ZNF69	55	GGGGTTGTTTGGATCGAT CGGAATC	56	CCCGCAAAAAAACCAA AATCTCGC
314	ZNF773	57	TAT AGG TTT CGA TGG CGG CGG TTA C	58	ATT CGA AAA AAA CCG AAA AAA CGC GAC
317	ZNF781	59	TAG TTT ATA AAC GCG GCG GAA TCG G	60	CCA ACG ACT ACT AAA TCA AAA AAC GCA

Each of the 29 MDMs (MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5) highly discriminated CC from benign cervico-vaginal (BCV) tissue with 10 MDMs having an area under the curve (AUC) >0.90 (Table VI). CC MDMs also highly discriminated in-situ adenocarcinomas (AIS) from BCV, but did not perform well in cervical intraepithelial neoplasia (CIN) 2/3 and CIN 1 (Table VI).

Table VI. Marker AUC values for discriminating CC from BCV, and positivity rates at 92.5% specificity for discriminating CC from BCV, AIS from BCV, CIN 2/3 from BCV, and CIN 1 from BCV.

Methylated DNA Marker (MDM)	AUC	Positivity Rates at 92.5% Specificity Cutoff in Controls			
	CC v. BCV	CC v BCV	AIS v. BCV	CIN 2/3 v. BCV	CIN 1 v. BCV
C1orf114	0.97 (0.94 - 1)	0.94 (0.86-0.98)	0.72 (0.55-0.86)	0.66 (0.47-0.81)	0.64 (0.31-0.89)
MAX.chr6.58147682-58147771	0.97 (0.93 - 1)	0.9 (0.81-0.96)	0.72 (0.55-0.86)	0.41 (0.24-0.59)	0.64 (0.31-0.89)
TTYH1	0.95 (0.91 - 1)	0.9 (0.81-0.96)	0.69 (0.52-0.84)	0.34 (0.19-0.53)	0.45 (0.17-0.77)
ZNF773	0.95 (0.91 - 0.99)	0.88 (0.78-0.94)	0.72 (0.55-0.86)	0.25 (0.11-0.43)	0.27 (0.06-0.61)
NEUROG3	0.95 (0.9 - 0.99)	0.88 (0.78-0.94)	0.72 (0.55-0.86)	0.25 (0.11-0.43)	0.45 (0.17-0.77)
ASCL1	0.94 (0.9 - 0.99)	0.91 (0.83-0.96)	0.69 (0.52-0.84)	0.34 (0.19-0.53)	0.18 (0.02-0.52)
NID2	0.92 (0.86 - 0.98)	0.8 (0.7-0.88)	0.56 (0.38-0.72)	0.22 (0.09-0.4)	0.27 (0.06-0.61)
ZNF781	0.91 (0.86 - 0.97)	0.85 (0.76-0.92)	0.64 (0.46-0.79)	0.31 (0.16-0.5)	0.27 (0.06-0.61)
CRHR2	0.91 (0.85 - 0.97)	0.85 (0.76-0.92)	0.67 (0.49-0.81)	0.38 (0.21-0.56)	0.27 (0.06-0.61)
MAX.chr9.36739811-36739868	0.9 (0.84 - 0.96)	0.79 (0.69-0.87)	0.58 (0.41-0.74)	0.38 (0.21-0.56)	0.45 (0.17-0.77)

From previous work it was recognized that the epigenetics of cancer subtypes within an organ differ and that the best panels are derived from combinations of subtype markers. The results are highlighted in Table VII. The discrimination strength of each marker assay is numerically denoted. A number of assays approached perfect discrimination for both adenocarcinomas and squamous cell cancers from benign tissues. Adenocarcinoma in-situ

samples were detected by several marker assays with AUCs in the high 80's and CIN2/3s in the low 80's. % methylation FC's for cancers vs controls ranged from 2 to 340.

Table VII. Marker AUC and fold change values discriminating adenocarcinoma from BCV, squamous cell cancer from BCV, and AIS from BCV.

Marker	AUC adenocarcinoma vs benign cervicovaginal (BCV)	Fold Change	AUC squamous cell cancer vs benign cervicovaginal	Fold Change	AUC adenocarcinoma in-situ vs benign cervicovaginal	AUC CIN 2/3 vs benign cervicovaginal
ABCB1	0.85	84	0.92	44	0.84	0.65
ARHGAP12	0.80	5	0.93	9	0.79	0.71
ASCL1	0.94	27	0.95	17	0.80	0.71
ATP10A (with Primer Seq ID Nos. 7 and 8)	0.73	9	0.88	13	0.57	0.54
ATP10A (with Primer Seq ID Nos. 9 and 10)	0.72	6	0.86	8	0.60	0.62
BARHL1	0.89	44	0.79	21	0.75	0.56
C1orf114	0.95	129	0.97	66	0.86	0.80
C2orf40	0.82	12	0.74	4	0.56	0.60
CACNA1C	0.93	27	0.84	12	0.83	0.38
CRHR2	0.91	46	0.88	20	0.82	0.57
HOPX_C	0.84	43	0.64	5	0.76	0.55
KCNQ5	0.82	10	0.56	2	0.70	0.54
MAX.chr1.98510968-98511049	0.91	81	0.79	23	0.82	0.72
MAX.chr18.73167751-73167791	0.85	18	0.86	12	0.73	0.71
MAX.chr2.127783183-127783403	0.90	34	0.85	13	0.76	0.52
MAX.chr4.8859853-8859939	0.91	15	0.80	8	0.78	0.60
MAX.chr6.58147682-58147771	0.97	24	0.97	16	0.85	0.83
MAX.chr9.36739811-36739868	0.90	15	0.90	17	0.74	0.65
NEUROG3	0.95	26	0.93	14	0.80	0.70
NID2	0.93	29	0.92	25	0.80	0.66
NXPH1	0.84	15	0.88	13	0.72	0.64
PRDM12	0.82	15	0.66	5	0.69	0.47
SLC9A3	0.86	26	0.71	7	0.68	0.59
ST8SIA1	0.93	42	0.81	18	0.81	0.54
TMEM200C	0.86	74	0.91	46	0.78	0.55
TTYH1	0.96	17	0.94	12	0.81	0.68
ZNF382	0.87	15	0.83	9	0.69	0.59
ZNF69	0.75	12	0.91	59	0.62	0.56
ZNF773	0.96	340	0.94	247	0.87	0.68
ZNF781	0.92	8	0.91	6	0.76	0.66

Experiments were also conducted that looked at the potential for CC MDM assays to separate gynecological malignancies. The organ site specificity of the assays was assessed which would be critical in a tampon based clinical test. Table VIII highlights the % methylation of 27 selected MDMs across the different cohorts tested. The 27 MDMs were highly methylated in the cervical cancer cohort, but generally < 5% in endometrial cancers.

Table VIII. Marker methylation AUC and fold change values discriminating cervical cancer from endometrial cancer.

Gene	Endometrial cancer methylation	Cervical Cancer Methylation	auc Cervical Cancer vs BCV	FC Cervical Cancer vs BCV	FC Cervical Cancer vs Endometrial Cancer
ABCB1	0.0170	0.7881	0.9062	225.30	46.49
C1orf95	0.0209	0.5044	0.9308	189.90	24.18
CACNA1C	0.0420	0.5008	0.9755	238.70	11.92
CACNG8	0.0177	0.3506	0.9273	62.21	19.78
CHST2	0.0050	0.5196	0.9022	280.80	104.44
ELMO1	0.0241	0.3677	0.9031	85.56	15.24
EMID2	0.0252	0.3303	0.9073	95.76	13.11
FBN1_B	0.0340	0.5594	0.9299	266.20	16.47
FLT3_A	0.0109	0.5418	0.9152	75.26	49.55
FLT3_B	0.0190	0.5145	0.9118	45.13	27.02
GLIS1	0.0196	0.4107	0.904	110.80	20.98
GPC6	0.0226	0.5852	0.9007	154.10	25.93
GREM2	0.0178	0.5089	0.9047	117.70	28.64
JAM2	0.0136	0.2773	0.9268	162.00	20.34
KCNK12_A	0.0365	0.5892	0.9048	202.20	16.12
LOC100129620	0.0463	0.8123	0.9062	290.70	17.55
LOC220930	0.0059	0.2983	0.924	120.90	50.49
MAX.chr15.78112404-78112692	0.0053	0.2422	0.9055	24.62	45.40
MAX.chr19.4584907-4585088	0.0298	0.57	0.9733	213.80	19.10
MAX.chr3.69591689-69591784	0.0260	0.5668	0.9256	85.32	21.80
NCAM1	0.0514	0.7085	0.9044	188.90	13.78
NT5C1A	0.0259	0.3969	0.9671	124.20	15.33
ST8SIA3	0.0239	0.5227	0.9273	88.40	21.85
ZNF382	0.0424	0.573	0.9152	126.80	13.51
ZNF419	0.0218	0.4464	0.9002	301.00	20.45
ZNF69	0.0618	0.4054	0.9377	56.69	6.56
ZSCAN18	0.0076	0.3602	0.9239	81.00	47.59

Whole methylome sequencing, stringent filtering criteria, and biological validation yielded outstanding candidate MDMs for cervical cancer. Some MDMs discriminate both CC

histologies from benign cervix with comparably high sensitivity, while others exhibit a subtype preference.

Samples:

Tissue and blood was obtained from Mayo Clinic biospecimen repositories with institutional IRB oversight. Samples were chosen with strict adherence to subject research authorization and inclusion/exclusion criteria. Cervical sub-types included 1) adenocarcinomas and, 2) squamous cell cancers. Controls included benign cervicovaginal (BCV) tissue and whole blood derived leukocytes. Endometrial cancers and controls were also run. Tissues were macro-dissected and histology reviewed by an expert GI pathologist. Samples were age sex matched, randomized, and blinded. DNA from 113 frozen tissues (16 grade 1/2 endometrioid (G1/2E), 16 grade 3 endometrioid (G3E), 11 serous, 11 clear cell ECs, 15 uterine carcinosarcomas, 44 benign endometrial (BE) tissues (14 proliferative, 12 atrophic, 18 disordered proliferative), 70 formalin fixed paraffin embedded (FFPE) cervical cancers (CC) (36 squamous cell, 34 adenocarcinomas), and 18 buffy coats from cancer-free females was purified using the QIAamp DNA Tissue Mini kit (frozen tissues), QIAamp DNA FFPE Tissue kit (FFPE tissues), and QIAamp DNA Blood Mini kit (buffy coat samples) (Qiagen, Valencia CA). DNA was re-purified with AMPure XP beads (Beckman-Coulter, Brea CA) and quantified by PicoGreen (Thermo-Fisher, Waltham MA). DNA integrity was assessed using qPCR.

Sequencing:

RRBS sequencing libraries were prepared following the Meissner protocol (Gu et al. Nature Protocols 2011) with modifications. Samples were combined in a 4-plex format and sequenced by the Mayo Genomics Facility on the Illumina HiSeq 2500 instrument (Illumina, San Diego CA). Reads were processed by Illumina pipeline modules for image analysis and base calling. Secondary analysis was performed using SAAP-RRBS, a Mayo developed bioinformatics suite. Briefly, reads were cleaned-up using Trim-Galore and aligned to the GRCh37/hg19 reference genome build with BSMAP. Methylation ratios were determined by calculating C/(C+T) or conversely, G/(G+A) for reads mapping to reverse strand, for CpGs with coverage $\geq$ 10X and base quality score $\geq$ 20.

Biomarker Selection:

Individual CpGs were ranked by hypermethylation ratio, namely the number of methylated cytosines at a given locus over the total cytosine count at that site. For cases, the ratios were required to be ≥ 0.20 (20%); for BCV tissue controls, ≤ 0.05 (5%); for buffy coat controls, ≤ 0.01 (1%). CpGs which did not meet these criteria were discarded. Subsequently, candidate CpGs were binned by genomic location into DMRs (differentially methylated regions) ranging from approximately 60 – 200bp with a minimum cut-off of 5 CpGs per region. DMRs with excessively high CpG density (>30%) were excluded to avoid GC-related amplification problems in the validation phase. For each candidate region, a 2-D matrix was created which compared individual CpGs in a sample to sample fashion for both cases and controls. Overall CC vs all benign endometria and/or no-cancer buffy coat were analyzed, as well as subtype comparisons. These CpG matrices were then compared back to the reference sequence to assess whether genomically contiguous methylation sites had been discarded during the initial filtering. From this subset of regions, final selections required coordinated and contiguous hypermethylation (in cases) of individual CpGs across the DMR sequence on a per sample level. Conversely, control samples had to have at least 10-fold less methylation than cases and the CpG pattern had to be more random and less coordinated. At least 10% of cancer samples within a subtype cohort were required to have at least a 50% hypermethylation ratio for every CpG site within the DMR.

In a separate analysis, a proprietary DMR identification pipeline and regression package was utilized to derive DMRs based on average methylation values of the CpG. The difference in average methylation percentage was compared between CC cases, tissue controls and buffy coat controls; a tiled reading frame within 100 base pairs of each mapped CpG was used to identify DMRs where control methylation was <5%; DMRs were only analyzed if the total depth of coverage was 10 reads per subject on average and the variance across subgroups was >0. Assuming a biologically relevant increase in the odds ratio of >3x and a coverage depth of 10 reads, ≥18 samples per group were required to achieve 80% power with a two-sided test at a significance level of 5% and assuming binomial variance inflation factor of one.

Following regression, DMRs were ranked by p-value, area under the receiver operating characteristic curve (AUC) and fold-change difference between cases and all controls. No adjustments for false discovery were made during this phase as independent validation was planned *a priori*.

Biomarker Validation:

A subset of the cervical cancer DMRs was chosen for further development. The criteria were primarily the logistic-derived area under the ROC curve metric which provides a performance assessment of the discriminant potential of the region. An AUC of 0.85 was chosen for the case vs control tissue comparison cut-off. 0.95 was the cut-off for the case vs blood comparison. In addition, the methylation fold-change ratio (average cancer hypermethylation ratio/average control hypermethylation ratio) was calculated and a lower limit of 20 was employed for tissue vs tissue comparisons and 50 for the tissue vs buffy coat comparisons. P values were required to be less than 0.05 and 0.001, respectively. DMRs had to be listed in both the average and individual CpG selection processes. Quantitative methylation specific PCR (qMSP) primers were designed for candidate regions using MethPrimer (see, Li LC and Dahiya R. Bioinformatics 2002 Nov;18(11):1427-31) and QC checked on 20ng (6250 equivalents) of positive and negative genomic methylation controls. Multiple annealing temperatures were tested for optimal discrimination. Validation was performed on a comparable set of independent tissue samples by qMSP. Additional cohorts included in-situ adenocarcinomas (AIS) and cervical intraepithelial neoplasia (CIN1-3). The patient demographics are shown in Table IX.

Table IX. Patient Demographics

	Benign (N=40)	CIN 1 (N=11)	CIN 2/3 (N=32)	Adenocarcinoma in situ (N=36)	Adenosquamous (N=1)	Squamous Cell (N=38)	Adenocarcinoma (N=43)
Age							
Mean (SD)	47.250 (14.143)	37.455 (8.251)	43.406 (14.771)	42.861 (10.139)	39.000 (NA)	48.053 (13.511)	42.000 (13.069)
Range	18.000 - 83.000	24.000 – 51.000	23.000 – 81.000	27.000 – 64.000	39.000 – 39.000	27.000 – 82.000	22.000 – 70.000
BMI							
N-Miss	0	1	1	4	0	6	13
Mean (SD)	28.295 (7.177)	26.581 (4.413)	27.138 (6.175)	28.718 (7.628)	37.090 (NA)	29.555 (11.873)	26.825 (6.387)
Range	17.640 – 53.600	22.030 – 35.520	19.230 – 43.170	17.720 – 45.200	37.090 – 37.090	19.530 – 86.760	18.980 – 41.980
Tobacco Use							
N-Miss	1	0	0	0	0	0	1

Never	23 (59.0%)	7 (63.6%)	18 (56.2%)	16 (44.4%)	0 (0.0%)	19 (50.0%)	24 (57.1%)
Ever	16 (41.0 %)	4 (36.4%)	14 (43.8%)	20 (55.6%)	1 (100.0%)	19 (50.0%)	18 (42.9%)
Number of pregnancies							
N-Miss	0	0	0	0	0	0	1
Mean (SD)	2.650 (1.494)	2.636 (2.461)	2.156 (2.302)	2.694 (1.600)	1.000 (NA)	2.447 (1.796)	1.857 (1.676)
Range	0.000 – 5.000	0.000 – 7.000	0.000 – 13.000	0.000 – 7.000	1.000 – 1.000	0.000 – 7.000	0.000 – 8.000
Number of live births							
Mean (SD)	2.200 (1.265)	2.000 (1.673)	1.750 (2.095)	2.167 (1.404)	1.000 (NA)	2.237 (1.715)	1.465 (1.202)
Range	0.000 – 5.000	0.000 – 4.000	0.000 – 12.000	0.000 – 5.000	1.000 – 1.000	0.000 – 7.000	0.000 – 5.000
Menopausal status							
Unknown	2 (5.0%)	1 (9.1%)	1 (3.1%)	6 (16.7%)	0 (0.0%)	8 (21.1%)	2 (4.7%)
Pre	23 (57.5%)	10 (90.9%)	22 (68.8%)	23 (63.9%)	1 (100.0%)	17 (44.7%)	30 (69.8%)
Peri	2 (5.0%)	0 (0.0%)	0 (0.0%)	2 (5.6%)	0 (0.0%)	2 (5.3%)	3 (7.0%)
Post	13 (32.5%)	0 (0.0%)	9 (28.1%)	5 (13.9%)	0 (0.0%)	11 (28.9%)	8 (18.6%)
HPV Status							
N-Miss	19	1	14	19	0	37	34
Unknown	2 (9.5%)	2 (20.0%)	6 (33.3%)	6 (35.3%)	1 (100.0%)	0 (0.0%)	5 (55.6%)
Negative	19 (90.5%)	1 (10.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Positive	0 (0.0%)	7 (70.0%)	12 (66.7%)	11 (64.7%)	0 (0.0%)	1 (100.0%)	4 (44.4%)
Insufficient	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Not completed	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

These tissues were identified as before, with expert clinical and pathological review. DNA purification was performed as previously described. The EZ-96 DNA Methylation kit

(Zymo Research, Irvine CA) was used for the bisulfite conversion step. 10ng of converted DNA (per marker) was amplified using SYBR Green detection on Roche 480 LightCyclers (Roche, Basel Switzerland). Serially diluted universal methylated genomic DNA (Zymo Research) was used as a quantitation standard. A CpG agnostic ACTB (β-actin) assay was used as an input reference and normalization control. Results were expressed as methylated copies (specific marker)/copies of ACTB.

Statistics:

Results were analyzed logistically for individual MDMs (methylated DNA marker) performance. For combinations of markers, random forest regression (rForest) was used to generated 500 individual models that were fit to boot strap samples of the original data (roughly 2/3 of the data for training) and used to estimate the cross-validation error (1/3 of the data for testing) of the entire MDM panel and was repeated 500 times. to avoid spurious splits that either under- or overestimate the true cross-validation metrics. Results were then averaged across the 500 iterations.

Example II.

This example describes the identification of methylated markers capable of distinguishing cervical cancer from endometrial cancer and ovarian cancer.

With the intent of uncovering organ specific hypermethylated regions among the three gynecological cancers (e.g., cervical, endometrial, and ovarian cancers) and their respective subtypes, experiments were conducted that combined or merged unique previously generated, validated, and disclosed tissue RRBS (reduced representation bisulfite sequencing) datasets for these cancers. Only those CpGs common to all three studies were analyzed. Regions had to contain at least 6 contiguous CpGs on either the sense or antisense strand. Discovery samples included 34 cervical adenocarcinomas, 36 cervical squamous cancers, 15 grade 1 or 2 uterine endometrioid cancers, 16 grade 3 uterine endometrioid cancers, 11 serous and 11 clear cell uterine cancers, 15 uterine carcinosarcomas, and 18 serous, 15 clear cell, 6 mucinous, and 18 endometrioid ovarian cancers. Benign controls included 18 cervical vaginal samples, 44 endometrial tissues (14 proliferative, 12 atrophic, 18 disordered proliferative, 10 secretory), 20 fallopian tube samples, and 36 non-cancer buffy coat or peripheral blood leukocyte samples.

For this application, experiments were conducted that focused on markers specific for cervical cancers – both adenocarcinoma and squamous cell varieties. As a first step, it was required that all regions of differential methylation had to have very low background or noise (<1%) in benign cervico-vaginal cells and tissues. DNA from this tissue type represents by far the highest proportion of nucleic acid found on a tampon. Second, so as to circumvent any potential signal from inflammation, regions with >1% methylation in leukocyte DNA were excluded. The remaining CpGs were used to compare cervical cancers to endometrial and ovarian cancers in aggregate. The adenocarcinoma and squamous cell cancers were analyzed separately. Three metrics were assessed: 1) the ratio of cervical cancer methylation to that for the endometrial and ovarian cancers; 2) the strength, frequency, and contiguous (read level) nature of the hypermethylation observed in the cervical cancer samples; and 3) the absolute methylation of the cervical cancers. For the first, cut-offs for CC/EC of > 20 and CC/OC >50 were used. These were chosen empirically to reduce the number of DMR candidates from the thousands to the hundreds. For the third, a cut-off of >20% was used. This reduced the number of candidates below 100. As for the second metric, a scoring system was utilized to identify concordant hypermethylation throughout the DMR. Regions with stochastic discordant CpG methylation were discarded. This analysis resulted in identification of a 64 DMR panel shown in Table X. Table XI shows the ratio of cervical cancer methylation to that for the endometrial cancer, ovarian cancer, and leukocyte (buffy coat) methylation for the markers recited in Table X.

Table X. Identified methylated regions distinguishing cervical cancer from endometrial cancer and ovarian cancer within tissue samples (the genomic coordinates for the regions shown are based on the Human Feb. 2009 (GRCh37/hg19) Assembly)

DMR	Gene Annotation	Chromosome No.	Chromosome Coordinates
7	AK5	1	77747411-77747907
362	BMP6	6	7726181-7726473
363	C12orf68	12	48577334-48577492
364	C13orf18	13	46961163-46961464
365	C1orf61	1	156391403-156391670
366	CHST10	2	101033659-101033898
367	COL13A1	10	71562414-71562559
368	COL19A1	6	70577142-70577563
369	DGKZ	11	46354636-46354919
370	EBF1	5	158526311-158526385
371	ELMOD1	11	107461821-107462200

372	EML6	2	54951792-54951889
373	FAM126A	7	23053556-23054025
374	FYN	6	112194444-112194670
375	GLT25D2	1	184006028-184006497
376	KCNA2	1	111149195-111149409
377	LOC100287216	2	109745183-109745250
378	LOC255130	4	58029916-58030017
379	MAST4	5	65892420-65892687
380	MAX.chr1.161582152-161582620	1	161582152-161582620
381	MAX.chr1.42501008-42501128	1	42501008-42501128
382	MAX.chr11.133920394-133920591	11	133920394-133920591
383	MAX.chr13.29394378-29394547	13	29394378-29394547
384	MAX.chr19.11805552-11805639	19	11805552-11805639
385	MAX.chr19.12098868-12099059	19	12098868-12099059
386	MAX.chr19.24216166-24216321	19	24216166-24216321
387	MAX.chr2.168149321-168149609	2	168149321-168149609
388	MAX.chr2.60808918-60809065	2	60808918-60809065
389	MAX.chr5.174220882-174220905	5	174220882-174220905
390	MAX.chr6.114663564-114663647	6	114663564-114663647
391	MAX.chr6.34113111-34113344	6	34113111-34113344
392	MAX.chr9.2242025-2242102	9	2242025-2242102
393	MAX.chr9.74061774-74061839	9	74061774-74061839
394	MED12L	3	150804540-150804760
395	MMP16	8	89339593-89339662
396	MYH10	17	8534612-8534792
397	NEGR1_B	1	72748492-72748610
398	NTNG1	1	107682964-107683280
399	PAQR9	3	142682557-142682820
400	PDE3B	11	14665532-14666436
401	PDE4A	19	10531731-10531820
402	PPM1E	17	56833450-56833580
403	PPP1R9A	7	94537822-94537957
404	RAB3C	5	57878796-57878920
405	SAMD5	6	147829302-147829357
406	SDC2	8	97506020-97506341
407	SDK2	17	71641381-71641605
408	SPINK2	4	57687751-57687933
409	TAF4B	18	23806366-23806832
410	TAF7	5	140700301-140700442
411	TRIM9	14	51561896-51562422
277	TRPC3_B	4	122872891-122873038
412	TSPAN5	4	99579208-99579526
413	ZNF14	19	19843701-19843789
414	ZNF211	19	58144584-58144700
415	ZNF280B	22	22862713-22862908
416	ZNF480	19	52800087-52800525
417	ZNF491	19	11909290-11909484
418	ZNF569	19	37960066-37960403
419	ZNF610	19	52839768-52839937

420	ZNF702P	19	53496769-53496864
421	ZNF709	19	12595655-12595874
422	ZNF845	19	53836815-53837093
423	ZNF91	19	23577939-23578063

Table XI. Ratio of cervical cancer methylation to that for the endometrial cancer (EC), ovarian cancer (OC), and leukocyte (buffy coat) methylation for the markers recited in Table X.

DMR	Gene Annotation	Ratio CC/EC	Ratio CC/OC	Ratio CC/Buffy
7	AK5	80	1178	277
362	BMP6	100	105	96
363	C12orf68	61	159	595
364	C13orf18	63	59	464
365	C1orf61	64	102	205
366	CHST10	173	193	859
367	COL13A1	55	90	177
368	COL19A1	82	97	169
369	DGKZ	42	53	136
370	EBF1	58	53	176
371	ELMOD1	105	115	266
372	EML6	52	99	374
373	FAM126A	126	483	234
374	FYN	118	683	562
375	GLT25D2	75	102	194
376	KCNA2	45	231	346
377	LOC100287216	156	204	645
378	LOC255130	220	>1000	116
379	MAST4	119	325	383
380	MAX.chr1.161582152-161582620	41	128	112
381	MAX.chr1.42501008-42501128	164	107	178
382	MAX.chr11.133920394-133920591	66	90	61
383	MAX.chr13.29394378-29394547	159	67	172
384	MAX.chr19.11805552-11805639	152	64	74
385	MAX.chr19.12098868-12099059	105	122	197
386	MAX.chr19.24216166-24216321	185	>1000	79
387	MAX.chr2.168149321-168149609	73	>1000	113
388	MAX.chr2.60808918-60809065	44	138	103
389	MAX.chr5.174220882-174220905	196	529	466
390	MAX.chr6.114663564-114663647	62	133	130
391	MAX.chr6.34113111-34113344	109	70	127
392	MAX.chr9.2242025-2242102	179	396	104
393	MAX.chr9.74061774-74061839	36	104	59
394	MED12L	113	>1000	367
395	MMP16	68	>1000	58
396	MYH10	112	367	229
397	NEGR1_B	64	62	184

398	NTNG1	50	108	225
399	PAQR9	81	180	190
400	PDE3B	120	174	711
401	PDE4A	90	59	268
402	PPM1E	80	923	279
403	PPP1R9A	37	102	38
404	RAB3C	70	302	238
405	SAMD5	93	76	110
406	SDC2	123	166	164
407	SDK2	46	>1000	217
408	SPINK2	99	69	758
409	TAF4B	244	1223	984
410	TAF7	68	193	58
411	TRIM9	145	74	175
277	TRPC3_B	156	813	95
412	TSPAN5	73	68	430
413	ZNF14	53	>1000	157
414	ZNF211	48	348	1776
415	ZNF280B	100	126	212
416	ZNF480	53	96	116
417	ZNF491	22	428	116
418	ZNF569	46	145	1418
419	ZNF610	40	278	52
420	ZNF702P	22	52	25
421	ZNF709	91	>1000	101
422	ZNF845	91	88	129
423	ZNF91	81	268	378

From these, due to DNA quantity limitations, the following eight DMRs were chosen to convert to qMSP (quantitative methylation specific PCR) assays and validate in an independent sample set: AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1. Primer sequences for the eight DMRs (AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1) are provided in Table XII.

Table XII. Primer sequences for AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1

DMR#	Name	SEQ ID NO:	Forward Primer 5'-3' Sequence (hg19)	SEQ ID NO:	Reverse Primer 5'-3' Sequence (hg19)
7	AK5	61	CGA CGT TTT ATT GCG TGC GTC GT	62	TTC CCT TAA CCA CCT AAT CCC CGA T
404	RABC3	63	GTT TGT CGG GAA TAT TCG GAG GGC	64	ACT ATC CTC TCC TAA CGC CGC ACA CG
417	ZNF491	65	TTA ATT CGG GGA AGT AGA AGG TCG T	66	AAA ACT AAA TAC AAA ACG CAA CGA A

419	ZNF610	67	ATT GAT TTA ACG TTT TGT TTC GCG T	68	AAA CGA AAT TAA AAA ACT CCC CGA A
423	ZNF91	69	CGG AGT TCG TTT GTT AAC GTA GTC GT	70	CCG AAT TCT CCT TAC CCA ACT CGA C
416	ZNF480	71	TAG ATT TCG GGT ATA GAA GCG CGG	72	ACA AAC CCG AAA ACG AAT CGC GTA
277	TRPC3_B	73	TTT CGC GGC GTT TTT TTA TTA TTT TTC GC	74	CTC CTA CCT TCC CGC CCT AAA CCG
371	ELMOD1	75	GCG GTT GTC GTA TTG GTT GC	76	GAA TAC ATC CCG ACT TAC TCC GCT

Samples included 38 cervical adenocarcinomas, 36 cervical squamous cell cancers, 18 grade 1 or 2 uterine endometrioid cancers, 24 grade 3 uterine endometrioid cancers, 16 serous and 7 clear cell uterine cancers, 18 uterine carcinosarcomas, and 36 serous, 21 clear cell, 4 mucinous, and 21 endometrioid ovarian cancers. Benign controls included 29 cervical vaginal samples, 14 endometrial tissues (8 proliferative, 2 atrophic, 3 disordered proliferative, 1 secretory), and 29 fallopian tube samples. The aggregate cervical cancers were compared logistically to the aggregate endometrial and ovarian cancers.

Individual MDMs (methylated DNA marker) ranged in performance from 30% sensitivity at 98% specificity to 73% sensitivity at 99% specificity. Panels of two to three MDMs were complementary. For example, AK5 and RABC3 together; 80% sensitivity at 98% specificity. Thus, marker combinations detected specific cervical cancer methylation 4/5s of the time with a 2% false positive rate.

In conclusion, these 8 MDMs (AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1), and the remaining 56 DMRs have the potential to indicate the presence of a cervical cancer, whether adenocarcinoma or squamous cell subtype, distinct from the two other gynecological organ cancers with high accuracy.

All publications and patents mentioned in the above specification are herein incorporated by reference in their entirety for all purposes. Various modifications and variations of the described compositions, methods, and uses of the technology will be apparent to those skilled in the art without departing from the scope and spirit of the technology as described. Although the technology has been described in connection with specific exemplary embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in

pharmacology, biochemistry, medical science, or related fields are intended to be within the scope of the following claims.

5

CLAIMS

WE CLAIM:

1. A method for characterizing a biological sample comprising:

measuring a methylation level of one or more methylated markers selected from Tables I, III, and X in the biological sample, wherein measuring a methylation level of one or more methylated markers comprises treating DNA from the biological sample with a reagent that modifies DNA in a methylation-specific manner.

2. The method of claim 1, wherein the biological sample is selected from a tissue sample, a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample, an organ secretion sample, a cerebrospinal fluid (CSF) sample, a saliva sample, a urine sample, and a stool sample.

3. The method of claim 2, wherein the tissue sample is a cervical tissue sample.

4. The method of claim 3, wherein the cervical tissue sample further comprises one or more of vaginal tissue, vaginal cells, endometrial tissue, endometrial cells, ovarian tissue, and ovarian cells.

5. The method of claim 2, wherein the secretion sample is a cervical secretion sample.

6. The method of claim 5, wherein the cervical secretion sample further comprises one or more of a vaginal secretion, an endometrial secretion, and an ovarian secretion.

7. The method of claim 1, wherein the measured methylation level of the one or more methylation markers is compared to a methylation level of a corresponding one or more methylation markers in control samples without cervical cancer.

8. The method of claim 7, further comprising determining that the individual has cervical cancer when the methylation level measured in the one or more methylation markers is higher than the methylation level measured in the respective control samples.

9. The method of claim 8, wherein the one or more methylated markers are selected from one of the following groups:

- the methylated markers recited in Tables I and/or III;
- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5;
- C1orf114, MAX.chr6.58147682-58147771, ZNF773, NEUROG3, ASCL1, NID2, ZNF781, CRHR2, and MAX.chr9.36739811-36739868; and
- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18.

10. The method of claim 5, further comprising determining that the individual has a subtype of cervical cancer.

11. The method of claim 10, wherein the subtype of cervical cancer is selected from cervical adenocarcinoma, and squamous cell cervical cancer.

12. The method of claim 10, wherein the one or more methylated markers are selected from one of the following groups:

- ABCB1, ARHGAP12, ASCL1, ATP10A, BARHL1, C1orf114, CACNA1C, CRHR2, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939,

MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, TMEM200C, TTYH1, ZNF382, ZNF69, ZNF773, and ZNF781.

13. The method of claim 7, further comprising determining that the individual has a cervical pre-cancer.

14. The method of claim 13, wherein the cervical pre-cancer is selected from cervix related in-situ adenocarcinoma, and cervical intraepithelial neoplasia.

15. The method of claim 13, wherein the one or more methylated markers are selected from one of the following groups:

- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ZNF773, TTYH1, NEUROG3, ZNF781, MAX.chr9.36739811-36739868, CRHR2, and NID2; and

ABCB1, ARHGAP12, ASCL1, ATP10A, BARHL1, C1orf114, CACNA1C, CRHR2, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, TMEM200C, TTYH1, ZNF382, ZNF69, ZNF773, and ZNF781.

16. The method of claim 1, wherein the measured methylation level of the one or more methylation markers is compared to a methylation level of a corresponding one or more methylation markers in endometrial cancer samples and/or ovarian cancer samples.

17. The method of claim 16, further comprising discriminating cervical cancer from endometrial cancer and/or ovarian cancer.

18. The method of claim 16, wherein the one or more methylated markers are selected from one of the following groups:

- the markers recited in Table X;
- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088,

MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18; and

- AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1.

19. The method of claim 1, wherein the reagent that modifies DNA in a methylation-specific manner is a borane reducing agent.

20. The method of claim 1, wherein the borane reducing agent is 2-picoline borane.

21. The method of claim 1, wherein the reagent that modifies DNA in a methylation-specific manner comprises one or more of a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent.

22. The method of claim 1, wherein the reagent that modifies DNA in a methylation-specific manner is a bisulfite reagent, and the treating produces bisulfite-treated DNA.

23. The method of claim 1, wherein the treated DNA is amplified with a set of primers specific for the one or more methylated markers.

24. The method of claim 23, wherein the set of primers specific for the one or more methylated markers is selected from the group recited in Tables V and XII.

25. The method of claim 23, wherein the set of primers specific for the one or more methylated markers is capable of binding an amplicon bound by a primer sequence for the specific methylated marker gene recited in Tables V and XII, wherein the amplicon bound by the primer sequence for the methylated marker gene recited in Tables V and XII is at least a portion of a genetic region for the methylated marker recited in Tables I, III, and X.

26. The method of claim 23, wherein the set of primers specific for the one or more methylated markers is a set of primers that specifically binds at least a portion of a

genetic region comprising chromosomal coordinates for a methylated marker recited in Tables I, III, and X.

27. The method of claim 1, wherein measuring a methylation level of one or more methylated markers comprises multiplex amplification.

28. The method of claim 1, wherein measuring a methylation level of one or more methylated markers comprises using one or more methods selected from the group consisting of methylation-specific PCR, quantitative methylation-specific PCR, methylation-specific DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, flap endonuclease assay, PCR-flap assay, and bisulfite genomic sequencing PCR.

29. The method of claim 1, wherein measuring a methylation level of one or more methylated markers comprises measuring methylation of a CpG site for the one or more methylation markers.

30. The method of claim 29, wherein the CpG site is present in a coding region or a regulatory region.

31. The method of claim 1, wherein the one or more methylated markers is described by the genomic coordinates shown in Tables I, III, and X.

32. The method of claim 1, wherein the biological sample is from a human subject.

33. The method of claim 32, wherein the human subject has or is suspected of having cervical cancer, a cervical cancer subtype, or a cervical pre-cancer.

34. The method of claim 2, wherein the biological sample is collected with a collection device having an absorbing member capable of collecting the biological sample upon contact with a bodily region.

35. The method of claim 34, wherein the absorbing member is a sponge having a shape and size suitable for insertion into a body orifice.

36. The method of claim 34, wherein the collection device is selected from a tampon, a lavage that releases liquid into the vagina and re-collects fluid, a cervical brush, a Fournier cervical self-sampling device, and a swab.

37. A method for preparing a deoxyribonucleic acid (DNA) fraction from a biological sample useful for analyzing one or more genetic loci involved in one or more chromosomal aberrations, comprising:

(a) extracting genomic DNA from a biological sample;

(b) producing a fraction of the extracted genomic DNA by:

(i) treating the extracted genomic DNA with a reagent that modifies DNA in a methylation-specific manner;

(ii) amplifying the treated genomic DNA using separate primers specific for one or more methylation markers recited in Tables I, III, and X;

(c) analyzing one or more genetic loci in the produced fraction of the extracted genomic DNA by measuring a methylation level for each of the one or more methylation markers.

38. The method of claim 37, wherein the reagent that modifies DNA in a methylation-specific manner is a borane reducing agent.

39. The method of claim 38, wherein the borane reducing agent is 2-picoline borane.

40. The method of claim 37, wherein the reagent that modifies DNA in a methylation-specific manner comprises one or more of a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent.

41. The method of claim 37, wherein the reagent that modifies DNA in a methylation-specific manner is a bisulfite reagent, and the treating produces bisulfite-treated DNA.

42. The method of claim 37, wherein the set of primers specific for the one or more methylated markers is selected from the group recited in Tables V and XII.

43. The method of claim 37, wherein the set of primers specific for the one or more methylated markers is capable of binding an amplicon bound by a primer sequence for the specific methylated marker gene recited in Tables V and XII, wherein the amplicon bound by the primer sequence for the methylated marker gene recited in Tables V and XII is at least a portion of a genetic region for the methylated marker recited in Tables I, III, and X.

44. The method of claim 37, wherein the set of primers specific for the one or more methylated markers is a set of primers that specifically binds at least a portion of a genetic region comprising chromosomal coordinates for a methylated marker recited in Tables I, III, and X.

45. The method of claim 37, wherein measuring a methylation level of one or more methylated markers comprises multiplex amplification.

46. The method of claim 37, wherein measuring a methylation level of one or more methylated markers comprises using one or more methods selected from the group consisting of methylation-specific PCR, quantitative methylation-specific PCR, methylation-specific DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, flap endonuclease assay, PCR-flap assay, and bisulfite genomic sequencing PCR.

47. The method of claim 37, wherein measuring a methylation level of one or more methylated markers comprises measuring methylation of a CpG site for the one or more methylation markers.

48. The method of claim 47, wherein the CpG site is present in a coding region or a regulatory region.

49. The method of claim 37, wherein the one or more methylated markers is described by the genomic coordinates shown in Tables I, III, and X.

50. The method of claim 37, wherein the biological sample is selected from a tissue sample, a blood sample, a plasma sample, a serum sample, a whole blood sample, a secretion sample, an organ secretion sample, a cerebrospinal fluid (CSF) sample, a saliva sample, a urine sample, and a stool sample.

51. The method of claim 50, wherein the tissue sample is a cervical tissue sample.

52. The method of claim 51, wherein the cervical tissue sample further comprises one or more of vaginal tissue, vaginal cells, endometrial tissue, endometrial cells, ovarian tissue, and ovarian cells.

53. The method of claim 50, wherein the secretion sample is a cervical secretion sample.

54. The method of claim 53, wherein the cervical secretion sample further comprises one or more of a vaginal secretion, an endometrial secretion, and an ovarian secretion.

55. The method of claim 37, wherein the biological sample is collected with a collection device having an absorbing member capable of collecting the biological sample upon contact with a bodily region.

56. The method of claim 55, wherein the absorbing member is a sponge having a shape and size suitable for insertion into a body orifice.

57. The method of claim 55, wherein the collection device is selected from a tampon, a lavage that releases liquid into the vagina and re-collects fluid, a cervical brush, a Fournier cervical self-sampling device, and a swab.

58. The method of claim 37, wherein the biological sample is from a human subject.

59. The method of claim 58, wherein the human subject has or is suspected of having cervical cancer, a cervical cancer subtype, or a cervical pre-cancer.

60. The method of claim 37, wherein the one or more methylated markers are selected from one of the following groups:

- MAX.chr6.58147682-58147771, C1ORF114, ASCL1, ARHGAP12, ZNF773, TTYH1, NEUROG3, ZNF781, NXPH1, MAX.chr9.36739811-36739868, NID2, TMEM200C, CRHR2, ABCB1, ZNF69, ATP10A, MAX.chr18.73167725-73167817, MAX.chr2.127783183-127783403, CACNA1C, ZNF382, BARHL1, MAX.chr4.8859853-8859939, ST8SIA1, MAX.chr1.98510958-98511049, C2ORF40, SLC9A3, PRDM12, HOPX_C, and KCNQ5;
- C1orf114, MAX.chr6.58147682-58147771, ZNF773, NEUROG3, ASCL1, NID2, ZNF781, CRHR2, and MAX.chr9.36739811-36739868;
- ABCB1, ARHGAP12, ASCL1, BARHL1, C1orf114, C2orf40, CACNA1C, CRHR2, HOPX_C, KCNQ5, MAX.chr1.98510968-98511049, MAX.chr18.73167751-73167791, MAX.chr2.127783183-127783403, MAX.chr4.8859853-8859939, MAX.chr6.58147682-58147771, MAX.chr9.36739811-36739868, NEUROG3, NID2, NXPH1, PRDM12, SLC9A3, TMEM200C, TTYH1, ZNF382, ZNF773, and ZNF781;
- ABCB1, c1orf95, CACNA1C, CACNG8, CHST2, ELMO1, EMID2, FBN1_B, FLT3_A, FLT3_B, GLIS1, GPC6, GREM2, JAM2, KCNK12_A, LOC100129620, MAX.chr15.78112404-78112692, MAX.chr19.4584907-4585088, MAX.chr3.69591689-69591784, NCAM1, NT5C1A, ST8SIA3, ZNF382, ZNF419, ZNF69, and ZSCAN18; and
- AK5, RABC3, ZNF491, ZNF610, ZNF91, ZNF480, TRPC3_B, and ELMOD1.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2022/019010

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 48/00; C12Q 1/00; C12Q 1/68; G01N 33/00 (2022.01)
CPC - C12Q 1/6886; C12Q 2600/106; C12Q 2600/118; C12Q 2600/154; C12Q 2600/156; C12Q 2600/16 (2022.05)

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)

see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	1st Invention:	
X	WO 2020/227100 A1 (CORNELL UNIVERSITY) 12 November 2020 (12.11.2020) entire document	1-3, 7-9, 16, 17, 21-23, 27-30, 32, 33
---		---
Y		4-6, 10, 11, 13, 14, 19, 20, 34-36
Y	US 2020/0377956 A1 (THE JOHNS HOPKINS UNIVERSITY et al) 03 December 2020 (03.12.2020) entire document	4-6, 10, 11
Y	*WO 2020/115728 A2 (HKG EPITHERAPEUTICS LIMITED) 11 June 2020 (11.06.2020) entire* document	*13, 14*
Y	US 2020/0024643 A1 (BLUESTAR GENOMICS INC. et al) 23 January 2020 (23.01.2020) entire document	19, 20
Y	US 2013/0337439 A1 (INFOGENE LDA. et al) 19 December 2013 (19.12.2013) entire document	34-36
A	US 2017/0314071 A1 (SEQUENOM INC.) 02 November 2017 (02.11.2017) entire document	1-11, 13, 14, 16, 17, 19-23, 27-30, 32-36

☒ 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
"D" document cited by the applicant in the international application
"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
27 June 2022	JUL 20 2022
Name and mailing address of the ISA/US *Mail Stop PCT, Attn: ISA/US, Commissioner for Patents* P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Taina Matos Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2022/019010

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	2nd Invention:	
X	WO 2020/227100 A1 (CORNELL UNIVERSITY) 12 November 2020 (12.11.2020) entire document	1-3, 7-9, 16, 17, 21-23, 27-30, 32, 33
---		---
Y		5, 6, 10, 11, 13, 14, 19, 20, 34-36
X	WO 2012/104642 A1 (UCL BUSINESS PLC) 09 August 2012 (09.08.2012) entire document	1-4
Y	US 2020/0377956 A1 (THE JOHNS HOPKINS UNIVERSITY et al) 03 December 2020 (03.12.2020) entire document	5, 6, 10, 11
Y	WO 2020/115728 A2 (HKG EPITHERAPEUTICS LIMITED) 11 June 2020 (11.06.2020) entire document	13, 14
Y	US 2020/0024643 A1 (BLUESTAR GENOMICS INC. et al) 23 January 2020 (23.01.2020) entire document	19, 20
Y	US 2013/0337439 A1 (INFOGENE LDA. et al) 19 December 2013 (19.12.2013) entire document	34-36
P,X	AHN et al., "Identification of novel DNA hypermethylation of the adenylate kinase 5 promoter in colorectal adenocarcinoma," Scientific Reports, 16 June 2021 (16.06.2021), Vol. 11: 12626, Pgs. 1-13. entire document	1-11, 13, 14, 16, 17, 19-23, 27-30, 32-36

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2022/019010

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2022/019010

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See extra sheet(s).

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

1st invention: 1-11, 13, 14, 16, 17, 19-23, 27-30, 32-36
2nd invention: 1-11, 13, 14, 16, 17, 19-23, 27-30, 32-36

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☒ No protest accompanied the payment of additional search fees.

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-60 are drawn to methods for characterizing a biological samples, and preparing a deoxyribonucleic acid (DNA) fraction from a biological sample.

The first invention of Group I+ is restricted to a methylated marker selected to be A1BG, and methods comprising the same. It is believed that claims 1-11, 13, 14, 16, 17, 19-23, 27-30, and 32-36 read on this first named invention and thus these claims will be searched without fee to the extent that they read on A1BG.

Applicant is invited to elect additional methylated markers, forward and reverse primers, and their respective, corresponding, SEQ ID NOs to be searched in a specific combination by paying additional fee for each set of election. An exemplary election would be a methylated marker selected to be ABCB1, a forward primer selected to be SEQ ID NO:1, a reverse primer selected to be SEQ ID NO:2. Additional methylated markers, forward and reverse primers, and their respective, corresponding, SEQ ID NOs will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The Groups I+ formulas do not share a significant structural element responsible for detecting cervical cancer, requiring the selection of alternative methylation markers where "A method for characterizing a biological sample comprising: measuring a methylation level of one or more methylated markers selected from Tables I, III, and X in the biological sample, wherein measuring a methylation level of one or more methylated markers comprises treating DNA from the biological sample with a reagent that modifies DNA in a methylation-specific manner."

Additionally, even if Groups I+ were considered to share the technical features of a method for characterizing a biological sample comprising: measuring a methylation level of one or more methylated markers in the biological sample, wherein measuring a methylation level of one or more methylated markers comprises treating DNA from the biological sample with a reagent that modifies DNA in a methylation-specific manner; a method for preparing a deoxyribonucleic acid (DNA) fraction from a biological sample useful for analyzing one or more genetic loci involved in one or more chromosomal aberrations, comprising: (a) extracting genomic DNA from a biological sample; (b) producing a fraction of the extracted genomic DNA by: (i) treating the extracted genomic DNA with a reagent that modifies DNA in a methylation-specific manner; (ii) amplifying the treated genomic DNA using separate primers specific for one or more methylation markers; (c) analyzing one or more genetic loci in the produced fraction of the extracted genomic DNA by measuring a methylation level for each of the one or more methylation markers. However, these shared technical features do not represent a contribution over the prior art.

Specifically, US 2017/0314071 A1 to Sequenom Inc. discloses a method for characterizing a biological sample (a method is provided for enriching fetal nucleic acids from a maternal biological sample, based on differential methylation, Para. [0007]; the state of methylation of a genomic sequence, refers to the characteristics of a DNA segment at a particular genomic locus relevant to methylation, Para. [0075]) comprising: measuring a methylation level of one or more methylated markers in the biological sample (quantification of fetal nucleic acid in a maternal sample based on the methylation status of the nucleic acid in said sample, Para. [0006]), wherein measuring a methylation level of one or more methylated markers comprises treating DNA from the biological sample with a reagent that modifies DNA in a methylation-specific manner (copy numbers obtained were calculated using the methylation markers and the Y-chromosome-specific markers using the assays provided in Table X, Para. [0060]; the sample is treated with a reagent that differentially modifies methylated and unmethylated DNA. For example, the reagent may comprise bisulfite; or the reagent may comprise one or more enzymes that preferentially cleave methylated DNA; or the reagent may comprise one or more enzymes that preferentially cleave unmethylated DNA, Para. [0020]); a method for preparing a deoxyribonucleic acid (DNA) fraction from a biological sample (preparing nucleic acid comprising a nucleotide sequence of the fetal nucleic acid by an amplification process in which fetal nucleic acid separated in part (b) is utilized as a template, Para. [0010]) useful for analyzing one or more genetic loci involved in one or more chromosomal aberrations (the amount of fetal nucleic acid may be determined at multiple loci across the genome, Para. [0040]), comprising: (a) extracting genomic DNA from a biological sample (the first steps of practicing the invention are to obtain a blood sample from a pregnant woman and extract DNA from the sample, Para. [0110]); (b) producing a fraction of the extracted genomic DNA (eluting the bound nucleic acid based on methylation status, wherein differentially methylated nucleic acids elute at least partly into separate fractions, Para. [0025]) by: (i) treating the extracted genomic DNA with a reagent that modifies DNA in a methylation-specific manner (copy numbers obtained were calculated using the methylation markers and the Y-chromosome-specific markers using the assays provided in Table X, Para. [0060]; the sample is treated with a reagent that differentially modifies methylated and unmethylated DNA. For example, the reagent may comprise bisulfite; or the reagent may comprise one or more enzymes that preferentially cleave methylated DNA; or the reagent may comprise one or more enzymes that preferentially cleave unmethylated DNA, Para. [0020]); (ii) amplifying the treated genomic DNA using separate primers specific for one or more methylation markers (designed to amplify nucleic acid greater than about 75, 80, 85, 90, 95, 100, 105, 110, 115 or 120 (or any number there between) base pairs thereby amplifying fetal nucleic acid and not digested maternal nucleic acid, Para. [0016]; PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels, Para. [0124]); (c) analyzing one or more genetic loci in the produced fraction of the extracted genomic DNA by measuring a methylation level for each of the one or more methylation markers (determining DNA methylation levels at specific gene loci in small amounts of genomic DNA ...[b]riefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA, Para. [0124]; the amount of fetal nucleic acid may be determined at multiple loci across the genome. Or when detecting the presence or absence of fetal aneuploidy, the amount of fetal nucleic acid may be determined at multiple loci on one or more target chromosomes, Para. [0040]).

International application No.
PCT/US2022/019010

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.