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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 January 2018 (11.01.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/009696 A1

(51) International Patent Classification:
C12N 15/11 (2006.01) *G01N 33/574* (2006.01)
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2017/040945

(22) International Filing Date:
06 July 2017 (06.07.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/358,776 06 July 2016 (06.07.2016) US

(71) Applicants: **YOUHEALTH BIOTECH, LIMITED** [GB/GB]; Suite #4-210, Governors Square, 23 Lime Tree Bay Avenue, KY1-1209 Grand Cayman (KY). **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, California 94607-5200 (US).

(72) Inventors: **ZHANG, Kang**; 14329 San Dieguito Road, San Diego, California 92130 (US). **HOU, Rui**; #17, Hunnan District, Pujiangyuan, Shenyang 110000 (CN). **ZHENG, Lianghong**; #17, Hunnan District, Pujiangyuan, Shenyang 110000 (CN).

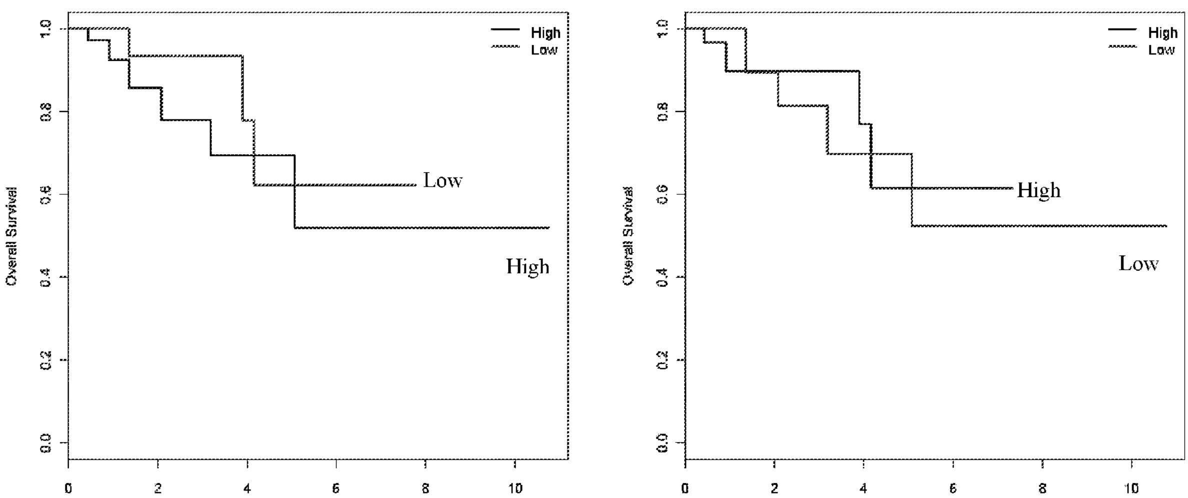
(74) Agent: **YANG, Xiaofan**; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, California 94304 (US).

(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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

(54) Title: COLON CANCER METHYLATION MARKERS AND USES THEREOF

Fig. 1



(57) Abstract: Disclosed herein are methods and kits for identifying a subject as having colon cancer. Also provided herein are methods and kits for determining the prognosis of a subject having colon cancer and for determining the progression of colon cancer in a subject.

[Continued on next page]

WO 2018/009696 A1

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

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))*

COLON CANCER METHYLATION MARKERS AND USES THEREOF

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 62/358,776, filed July 6, 2016, which application is incorporated herein by reference.

BACKGROUND OF THE DISCLOSURE

[0002] Cancer is a leading cause of deaths worldwide, with annual cases expected to increase from 14 million in 2012 to 22 million during the next two decades (WHO). Diagnostic procedures for colon cancer, in some cases, begin only after a patient is already present with symptoms, leading to costly, invasive, and sometimes time-consuming procedures. In addition, inaccessible areas sometimes prevent an accurate diagnosis. Further, high cancer morbidities and mortalities are associated with late diagnosis.

SUMMARY OF THE DISCLOSURE

[0003] Provided herein are methods and kits for identifying a subject as having colon cancer. Also provided herein are methods and kits for determining the prognosis of a subject having colon cancer. Further provided herein are methods and kits for determining the progression of colon cancer in a subject.

[0004] In certain embodiments, provided herein is a method of selecting a subject suspected of having colon cancer for treatment, the method comprising: (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having colon cancer; (b) generating a methylation profile comprising one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24583770, cg24741563, cg22979615, and cg10673833 from the extracted genomic DNA; (c) comparing the methylation profile of the one or more biomarkers with a control; (d) identifying the subject as having colon cancer if the methylation profile correlates to the control; and (e) administering an effective amount of a therapeutic agent to the subject if the subject is identified as having colon cancer.

[0005] In some embodiments, the methylation profile comprises cg10673833.

[0006] In some embodiments, the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24583770, cg24741563, and cg22979615. In some embodiments, the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615.

[0007] In some embodiments, the methylation profile comprises cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615.

[0008] In some embodiments, the comparing further comprises generating a pair-wise methylation difference dataset comprising: (i) a first difference between the methylation profile of the treated genomic DNA with a methylation profile of a first normal sample; (ii) a second difference between a methylation profile of a second normal sample and a methylation profile of a third normal sample; and (iii) a third difference between a methylation profile of a first primary cancer sample and a methylation profile of a second primary cancer sample.

[0009] In some embodiments, the comparing further comprises analyzing the pair-wise methylation difference dataset with a control by a machine learning method to generate the methylation profile.

[0010] In some embodiments, the first primary cancer sample is a colon cancer sample.

[0011] In some embodiments, the second primary cancer sample is a non-colon cancer sample.

[0012] In some embodiments, the control comprises a set of methylation profiles, wherein each said methylation profile is generated from a biological sample obtained from a known cancer type.

[0013] In some embodiments, the known cancer type is colon cancer.

[0014] In some embodiments, the known cancer type is a relapsed or refractory colon cancer.

[0015] In some embodiments, the known cancer type is a metastatic colon cancer.

[0016] In some embodiments, the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.

[0017] In some embodiments, the generating further comprises hybridizing each of the one or more biomarkers with a probe, and performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers.

[0018] In some embodiments, the biological sample comprises a blood sample. In some embodiments, the biological sample comprises a tissue biopsy sample. In some embodiments, the biological sample comprises circulating tumor cells.

[0019] In some embodiments, the subject is a human.

[0020] In certain embodiments, provided herein is a method of generating a methylation profile of a biomarker in a subject in need thereof, comprising: (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject; (b) detecting a hybridization between the extracted genomic DNA and a probe, wherein the probe hybridizes to a biomarker selected from cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24583770, cg24741563, cg22979615, and cg10673833; and (c) generating a methylation profile based on the detected hybridization between the extracted genomic DNA and the probe.

[0021] In some embodiments, the methylation profile comprises cg10673833.

[0022] In some embodiments, the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24583770, cg24741563, and cg22979615. In some embodiments, the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615.

[0023]

[0024] In some embodiments, the methylation profile comprises cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615.

[0025] In some embodiments, the generating further comprises generating a pair-wise methylation difference dataset comprising: (i) a first difference between the methylation profile of the treated genomic DNA with a methylation profile of a first normal sample; (ii) a second difference between a methylation profile of a second normal sample and a methylation profile of a third normal sample; and (iii) a third difference between a methylation profile of a first primary cancer sample and a methylation profile of a second primary cancer sample.

[0026] In some embodiments, the generating further comprises analyzing the pair-wise methylation difference dataset with a control by a machine learning method to generate the methylation profile.

[0027] In some embodiments, the first primary cancer sample is a colon cancer sample.

[0028] In some embodiments, the second primary cancer sample is a non-colon cancer sample.

[0029] In some embodiments, the control comprises a set of methylation profiles, wherein each said methylation profile is generated from a biological sample obtained from a known cancer type.

[0030] In some embodiments, the known cancer type is colon cancer.

[0031] In some embodiments, the known cancer type is a relapsed or refractory colon cancer.

[0032] In some embodiments, the known cancer type is a metastatic colon cancer.

[0033] In some embodiments, the known cancer type is adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma.

[0034] In some embodiments, the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.

[0035] In some embodiments, the method further comprises performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers prior to generating the methylation profile.

[0036] In some embodiments, the biological sample comprises a blood sample. In some embodiments, the biological sample comprises a tissue biopsy sample. In some embodiments, the biological sample comprises circulating tumor cells.

[0037] In some embodiments, the subject is a human.

[0038] In certain embodiments, provided herein is a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in the subject, comprising: (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject having colon cancer; (b) generating a methylation profile comprising one or more biomarkers selected from: cg13510262, cg13859324, cg21685565 and cg26952662 from the extracted genomic DNA; (c) obtaining a methylation score based on the methylation profile of the one or more biomarkers; and (d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission, initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent.

[0039] In some embodiments, the methylation profile comprises cg13510262, cg13859324, and cg26952662.

[0040] In some embodiments, the methylation profile further comprises one or more biomarkers selected from: cg14159672, cg17675150, cg13620770, and cg21922841.

[0041] In some embodiments, the methylation profile further comprises one or more biomarkers selected from: cg13620770 and cg21922841.

[0042] In some embodiments, the methylation profile further comprises one or more biomarkers selected from: cg14159672 and cg17675150.

[0043] In some embodiments, the methylation profile further comprises cg14159672, cg17675150, cg13620770, and cg21922841.

[0044] In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 6 months. In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 1 year. In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 1.5 years. In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 2 years. In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 2.5 years. In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 3 years. In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 4 years. In some embodiments, the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 5 years.

[0045] In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 6 months. In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 1 year. In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 1.5 years. In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 2 years. In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 2.5 years. In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 3 years. In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 4 years. In some embodiments, the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 5 years.

[0046] In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 5 years. In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 4 years. In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 3 years. In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 2.5 years. In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 2 years. In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 1.5 years. In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 1 year. In some embodiments, the methylation score of less than 1.5 is indicative of a survival of less than 6 months.

[0047] In some embodiments, the methylation score is calculated based on Cox proportional hazards (PH) regression analysis.

[0048] In some embodiments, colon cancer is metastatic colon cancer.

[0049] In some embodiments, colon cancer is adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma.

[0050] In some embodiments, the generating further comprises hybridizing each of the one or more biomarkers with a probe, and performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers.

[0051] In some embodiments, the biological sample comprises a blood sample. In some embodiments, the biological sample comprises a tissue biopsy sample. In some embodiments, the biological sample comprises circulating tumor cells.

[0052] In some embodiments, the subject is a human.

[0053] In certain embodiments, provided herein is a kit comprising a set of nucleic acid probes that hybridizes to biomarkers: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the set of nucleic acid probes comprises a set of padlock probes.

[0054] In certain embodiments, provided herein is a kit comprising a set of nucleic acid probes that hybridizes to biomarkers: cg13510262, cg13859324, and cg26952662. In some embodiments, the kit further comprises a nucleic acid probe that hybridizes to a biomarker selected from cg14159672, cg17675150, cg13620770, and cg21922841. In some embodiments, the set of nucleic acid probes comprises a set of padlock probes.

BRIEF DESCRIPTION OF THE DRAWINGS

[0055] Various aspects of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0056] Fig. 1 illustrates Kaplan-Meier curves from LASSO (left panel) and Bosting (right panel).

[0057] Fig. 2 shows Cox proportional hazards regression prediction curves from LASSO (left panel) and Boosting (right panel).

[0058] Fig. 3 shows a multiclass ROC curve from LASSO.

[0059] Fig. 4 shows a boxplot for methylation change in different response groups for colon cancer.

[0060] Fig. 5 shows a boxplot for methylation change in different response groups for colon cancer.

[0061] Fig. 6 illustrates an exemplary BCM for colon cancer.

DETAILED DESCRIPTION OF THE DISCLOSURE

[0062] Cancer is characterized by an abnormal growth of a cell caused by one or more mutations or modifications of a gene leading to dysregulated balance of cell proliferation and cell death. DNA methylation silences expression of tumor suppression genes, and presents itself as one of the first neoplastic changes. Methylation patterns found in neoplastic tissue and plasma demonstrate homogeneity, and in some instances are utilized as a sensitive diagnostic marker. For example, cMethDNA assay has been shown in one study to be about 91% sensitive and about 96% specific when used to diagnose metastatic breast cancer. In another study, circulating tumor DNA (ctDNA) was about 87.2% sensitive and about 99.2% specific when it was used to identify KRAS gene mutation in a large cohort of patients with metastatic colon cancer (Bettegowda et al., Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Sci. Transl. Med, 6(224):ra24. 2014). The same study further demonstrated that ctDNA is detectable in >75% of patients with advanced pancreatic, ovarian, colorectal, bladder, gastroesophageal, breast, melanoma, hepatocellular, and head and neck cancers (Bettegowda et al).

[0063] Additional studies have demonstrated that CpG methylation pattern correlates with neoplastic progression. For example, in one study of breast cancer methylation patterns, P16 hypermethylation has been found to correlate with early stage breast cancer, while TIMP3 promoter hypermethylation has been correlated with late stage breast cancer. In addition, BMP6, CST6 and TIMP3 promoter hypermethylation have been shown to associate with metastasis into lymph nodes in breast cancer.

[0064] In some embodiments, DNA methylation profiling provides higher clinical sensitivity and dynamic range compared to somatic mutation analysis for cancer detection. In other instances, altered DNA methylation signature has been shown to correlate with the prognosis of treatment response for certain cancers. For example, one study illustrated that RASSF1A DNA methylation measurement in serum was used to predict a poor outcome in patients undergoing adjuvant therapy in breast cancer patients. In addition, another study has demonstrated that ESR1 gene methylation correlate with clinical response in breast cancer patients receiving tamoxifen. Additionally, ARHI gene promoter hypermethylation was shown to be a predictor of long-term survival in breast cancer patients not treated with tamoxifen.

[0065] In some embodiments, disclosed herein are methods and kits of diagnosing colon cancer based on DNA methylation profiling. In some instances, provided herein are methods and kits of identifying a subject has having colon cancer based on the DNA methylation profiling. In some instances, also provided herein are methods and kits of determining the prognosis of a subject having colon cancer and determining the progression of colon cancer in a subject based on the DNA methylation profilings.

Methods of Use

Methods of Diagnosis of a Subject

[0066] Disclosed herein, in certain embodiments, are methods of diagnosing colon cancer and selecting subjects suspected of having colon cancer for treatment. In some instances, the methods comprise utilizing one or more biomarkers described herein. In some instances, a biomarker comprises a cytosine methylation site. In some instances, cytosine methylation comprises 5-methylcytosine (5-mCyt) and 5-hydroxymethylcytosine. In some cases, a cytosine methylation site occurs in a CpG dinucleotide motif. In other cases, a cytosine methylation site occurs in a CHG or CHH motif, in which H is adenine, cytosine or thymine. In some instances, one or more CpG dinucleotide motif or CpG site forms a CpG island, a short DNA sequence rich in CpG dinucleotide. In some instances, CpG islands are typically, but not always, between about 0.2 to about 1 kb in length. In some instances, a biomarker comprises a CpG island.

[0067] In some embodiments, disclosed herein is a method of selecting a subject suspected of having colon cancer for treatment, in which the method comprises (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having colon cancer; (b) generating a methylation profile comprising one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24583770, cg24741563, cg22979615, and cg10673833 from the extracted genomic DNA; (c) comparing the methylation profile of the one or more biomarkers with a control; (d) identifying the subject as having colon cancer if the methylation profile correlates to the control; and (e) administering an effective amount of a therapeutic agent to the subject if the subject is identified as having colon cancer.

[0068] In some embodiments, a methylation profile comprises a plurality of CpG methylation data for one or more biomarkers described herein. In some instances, a plurality of CpG methylation data is generated by first obtaining a genomic DNA (e.g., nuclear DNA or circulating DNA) from a biological sample, and then treating the genomic DNA by a deaminating agent to generate an extracted genomic DNA. In some instances, the extracted genomic DNA (e.g., extracted nuclear DNA or extracted circulating DNA) is optionally treated with one or more restriction enzymes to generate a set of DNA fragments prior to submitting for sequencing analysis to generate CpG methylation data. In some cases, the sequencing analysis comprises hybridizing each of the one or more biomarkers described herein with a probe, and performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers. In some instances, the CpG methylation data is then input into a machine learning/classification program to generate a methylation profile.

[0069] In some instances, a set of biological samples are generated and subsequently input into the machine learning/classification program. In some instances, the set of biological samples comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, or more biological samples. In some instances, the set of biological samples comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, or more normal biological samples. In some instances, the set of biological samples comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, or more cancerous biological samples. In some cases, the set of biological samples comprise a biological sample of interest, a first primary cancer sample, a second primary cancer sample, a first normal sample, a second normal sample, and a third normal sample; wherein the first, and second primary cancer samples are different; and wherein the first, second, and third normal samples are different. In some cases, three pairs of difference datasets are generated in which the three pairs of dataset comprise: a first difference dataset between the methylation profile of the biological sample of interest and the first normal sample, in which the biological sample of interest and the first normal sample are from the same biological sample source; a second difference dataset between a methylation profile of a second normal sample and a methylation profile of a third normal sample, in which the second and third normal samples are different; and a third difference dataset between a methylation profile of a first primary cancer sample and a methylation profile of a second primary cancer sample, in which the first and second primary cancer samples are different. In some instances, the difference datasets are further input into the machine learning/classification program. In some cases, a pair-wise methylation difference dataset from the first, second, and third datasets is generated and then analyzed in the presence of a control dataset or a training dataset by the machine learning/classification method to generate the cancer CpG methylation profile. In some instances, the first primary cancer sample is a colon cancer sample. In some cases, the second primary cancer sample is a non-colon cancer sample. In some cases, the machine learning method comprises identifying a plurality of markers and a plurality of weights based on a top score (e.g., a t-test value, a β test value), and classifying the samples based on the plurality of markers and the plurality of weights. In some cases, the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.

[0070] In some embodiments, the CpG methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24583770, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises two or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises three or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises four or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises five or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises six or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises seven or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises eight or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833. In some embodiments, the CpG methylation profile comprises cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833.

[0071] In some instances, the CpG methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615. In some instances, the CpG methylation profile comprises two or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615. In some instances, the CpG methylation profile comprises three or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615. In some instances, the CpG methylation profile comprises four or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615. In some instances, the CpG methylation profile comprises five or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615. In some instances, the CpG methylation profile comprises six or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615. In some instances, the CpG methylation profile comprises seven or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615. In some instances, the CpG methylation profile comprises cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, and cg22979615.

[0072] In some instances, the CpG methylation profile comprises biomarker cg10673833.

[0073] In some instances, the subject is diagnosed in having colon cancer. In some instances, colon cancer further comprises a relapsed or refractory colon cancer. In other instances, colon cancer comprises a metastatic colon cancer. In some cases, the subject is diagnosed in having a relapsed or refractory colon cancer. In additional cases, the subject is diagnosed in having a metastatic colon cancer.

[0074] In some embodiments, a colon cancer is any type of colon cancer. In some instances, colon cancer also refers to colorectal cancer. In some instances, a colon cancer comprises adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma.

[0075] In some embodiments, the subject diagnosed of having colon cancer is further treated with a therapeutic agent. Exemplary therapeutic agents include, but are not limited to, bevacizumab, capecitabine, cetuximab, doxorubicin, fluorouracil injection, irinotecan hydrochloride, leucovorin calcium, oxaliplatin, panitumumab, ramucirumab, regorafenib, Ziv-aflibercept, or a combination thereof.

[0076] In some embodiments, also described herein include a method of generating a methylation profile of a biomarker. In some instances, the method comprises (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject; (b) detecting a hybridization between the extracted genomic DNA and a probe, wherein the probe hybridizes to a biomarker selected from cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24583770, cg24741563, cg22979615, and cg10673833; and (c) generating a methylation profile based on the detected hybridization between the extracted genomic DNA and the probe.

[0077] In some embodiments, one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833 are used to generate a methylation profile. In some embodiments, two or more, three or more, four or more, five or more, six or more, seven or more, or eight or more, biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833 are used to generate the methylation profile. In some embodiments, cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833 are used to generate a methylation profile.

[0078] In some instances, as described elsewhere herein, a pair-wise methylation difference dataset is generated prior to generating a methylation profile. In some cases, the pair-wise methylation difference dataset comprises (i) a first difference between the methylation profile of the treated genomic DNA with a methylation profile of a first normal sample; (ii) a second difference between a methylation profile of a second normal sample and a methylation profile of a third normal sample; and (iii) a third difference between a methylation profile of a first primary cancer sample and a methylation profile of a second primary cancer sample.

[0079] In some cases, the pair-wise methylation difference dataset is analyzed with a control by a machine learning method to generate a methylation profile. In some cases, the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.

[0080] In some embodiments, a probe comprises a DNA probe, RNA probe, or a combination thereof. In some instances, a probe comprises natural nucleic acid molecules and non-natural nucleic acid molecules. In some cases, a probe comprises a labeled probe, such as for example, fluorescently labeled probe or radioactively labeled probe. In some instances, a probe correlates to a CpG site. In some instances, a probe is utilized in a next generation sequencing reaction to generate a CpG methylation data. In further instances, a probe is used in a solution-based next generation sequencing reaction to generate a CpG methylation data. In some cases, a probe comprises a molecular beacon probe, a TaqMan probe, locked nucleic acid probe, a pad-lock probe, or Scorpion probe. In some cases, a probe comprises a pad-lock probe.

[0081] In some cases, the method further comprises performing a DNA sequencing reaction such as those described elsewhere herein to quantify the methylation of each of the one or more biomarkers prior to generating a methylation profile.

[0082] In some embodiments, a CpG methylation site is located at the promoter region (e.g., induces a promoter methylation). In some instances, promoter methylation leads to a downregulation of its corresponding gene expression. In some instances, one or more CpG methylation sites described *supra* and in subsequent paragraphs are located at promoter regions, leading to promoter methylation, and subsequent downregulation of the corresponding gene expression. In some instances, the CpG methylation site is as illustrated in Tables 11 (e.g., Table 11A) or 12. In some cases, an increase in gene expression leads to a decrease in tumor volume.

[0083] In some embodiments, one or more cg markers reference one or more genes. In some embodiments, cg08088171 references calcium activated nucleotidase 1 (*CANT1*). In some embodiments, cg27536151 references mitogen-activated protein kinase kinase 5 (*MAP2K5*). In some embodiments, cg13420112 references zinc finger protein 512B (*ZNF512B*). In some embodiments, cg14642259 references myosin binding protein C, cardiac (*MYBPC3*). In some embodiments, cg20973720 references Ras protein specific guanine nucleotide releasing factor 1 (*RASGRF1*). In some embodiments, cg24741563 references amyloid beta precursor protein binding protein 2 (*APPBP2*). In some embodiments, cg22979615 references MOB kinase activator 2 (*HCCA2*). In some embodiments, cg10673833 references myosin IG (*MYO1G*).

[0084] In some embodiments, described herein is a method of selecting a subject suspected of having colon cancer for treatment, the method comprising generating a methylation profile comprising one or more genes selected from: calcium activated nucleotidase 1 (*CANT1*), mitogen-activated protein kinase kinase 5 (*MAP2K5*), zinc finger protein 512B (*ZNF512B*), myosin binding protein C, cardiac (*MYBPC3*), Ras protein specific guanine nucleotide releasing factor 1 (*RASGRF1*), amyloid beta precursor protein binding protein 2 (*APPBP2*), MOB kinase activator 2 (*HCCA2*), and myosin IG (*MYO1G*). In some embodiments, the methylation profile comprises one or more genes selected from: calcium activated nucleotidase 1 (*CANT1*), mitogen-activated protein kinase kinase 5 (*MAP2K5*), zinc finger protein 512B (*ZNF512B*), myosin binding protein C, cardiac (*MYBPC3*), Ras protein specific guanine nucleotide releasing factor 1 (*RASGRF1*), amyloid beta precursor protein binding protein 2 (*APPBP2*), and (MOB kinase activator 2 (*HCCA2*). In some embodiments, the methylation profile comprises calcium activated nucleotidase 1 (*CANT1*), mitogen-activated protein kinase kinase 5 (*MAP2K5*), zinc finger protein 512B (*ZNF512B*), myosin binding protein C, cardiac (*MYBPC3*), Ras protein specific guanine nucleotide releasing factor 1 (*RASGRF1*), amyloid beta precursor protein binding protein 2 (*APPBP2*), and MOB kinase activator 2 (*HCCA2*). In some embodiments, the methylation profile comprises myosin IG (*MYO1G*).

[0085] In some embodiments, described herein is a method of generating a methylation profile of a biomarker in a subject in need thereof, comprising detecting a hybridization between the extracted genomic DNA and a probe, in which the probe hybridizes to a gene selected from: calcium activated nucleotidase 1 (*CANT1*), mitogen-activated protein kinase kinase 5 (*MAP2K5*), zinc finger protein 512B (*ZNF512B*), myosin binding protein C, cardiac (*MYBPC3*), Ras protein specific guanine nucleotide releasing factor 1 (*RASGRF1*), amyloid beta precursor protein binding protein 2 (*APPBP2*), MOB kinase activator 2 (*HCCA2*), and myosin IG (*MYO1G*). In some embodiments, the probe hybridizes to a gene selected from: calcium activated nucleotidase 1 (*CANT1*), mitogen-activated protein kinase kinase 5 (*MAP2K5*), zinc finger protein 512B (*ZNF512B*), myosin binding protein C, cardiac (*MYBPC3*), Ras protein specific guanine nucleotide releasing factor 1 (*RASGRF1*), amyloid beta precursor protein binding protein 2 (*APPBP2*), and MOB kinase activator 2 (*HCCA2*). In some embodiments, the probe hybridizes myosin IG (*MYO1G*).

Determining the Prognosis of a Subject Having Colon Cancer or Monitoring the Progression of Colon Cancer in a Subject

[0086] In some embodiments, disclosed herein include a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in a subject. In some instances, colon cancer comprises adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma. In some instances, disclosed herein is a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in a subject. In some embodiments, the method comprises (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject having colon cancer; (b) generating a methylation profile comprising one or more biomarkers selected from: cg13510262, cg13859324, cg21685565, and cg26952662 from the extracted genomic DNA; (c) obtaining a methylation score based on the methylation profile of the one or more biomarkers; and (d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission, initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent.

[0087] In some instances, the methylation profile comprises two or more biomarkers selected from: cg13510262, cg13859324, and cg26952662. In some instances, the methylation profile comprises three or more biomarkers selected from: cg13510262, cg13859324, and cg26952662. In some instances, the methylation profile comprises cg13510262, cg13859324, and cg26952662.

[0088] In some cases, the methylation profile further comprises one or more biomarkers selected from: cg14159672, cg17675150, cg13620770, and cg21922841. In some cases, the methylation profile further comprises two or more biomarkers selected from: cg14159672, cg17675150, cg13620770, and cg21922841. In some cases, the methylation profile further comprises three or more biomarkers selected from: cg14159672, cg17675150, cg13620770, and cg21922841. In some cases, the methylation profile further comprises biomarkers cg14159672, cg17675150, cg13620770, and cg21922841.

[0089] In some cases, the methylation profile further comprises one or more biomarkers selected from: cg13620770 and cg21922841.

[0090] In some cases, the methylation profile further comprises one or more biomarkers selected from: cg14159672 and cg17675150.

[0091] In some instances, the methylation profile comprises one or more biomarkers selected from: cg13510262, cg13859324, and cg26952662; and one or more biomarkers selected from: cg14159672, cg17675150, cg13620770, and cg21922841.

[0092] In some embodiments, one or more cg markers reference one or more genes. In some embodiments, cg13510262 references Sp1 transcription factor (*SP1*). In some embodiments, cg13859324 references Unc-45 myosin chaperone B (*UNC45B*). In some embodiments, cg26952662 references collagen triple helix repeat containing 1 (*CTHRC1*). In some embodiments, cg17675150 references zinc finger protein 532 (*ZNF532*). In some embodiments, cg13620770 references G protein-coupled receptor 137 (*GPR137*). In some embodiments, cg21922841 references SLC9A3 regulator 1 (*SLC9A3R1*).

[0093] In some embodiments, described herein is a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in the subject, comprising generating a methylation profile comprising one or more genes selected from: Sp1 transcription factor (*SP1*), Unc-45 myosin chaperone B (*UNC45B*), and collagen triple helix repeat containing 1 (*CTHRC1*). In some embodiments, the methylation profile further comprises one or more genes selected from: zinc finger protein 532 (*ZNF532*), G protein-coupled receptor 137 (*GPR137*), and SLC9A3 regulator 1 (*SLC9A3R1*). In some embodiments, the methylation profile further comprises peptidase M20 domain containing 1 (*PM20D1*), zinc finger protein 532 (*ZNF532*), G protein-coupled receptor 137 (*GPR137*), and SLC9A3 regulator 1 (*SLC9A3R1*). In some embodiments, the methylation profile further comprises G protein-coupled receptor 137 (*GPR137*) and SLC9A3 regulator 1 (*SLC9A3R1*). In some embodiments, the methylation profile further comprises peptidase M20 domain containing 1 (*PM20D1*) and zinc finger protein 532 (*ZNF532*).

Methylation Scores

[0094] In some instances, a methylation score is utilized to determine the prognosis of a subject. In some instances, prognosis refers to the prediction of the likelihood of cancer-attributable death or progression, including recurrence, metastatic spread, and drug resistance, of colon cancer. The term "prediction" is used herein to refer to the likelihood that a subject will respond either favorably or unfavorably to a drug or set of drugs, and also the extent of those responses, or that a subject will survive, following chemotherapy for a certain period of time without cancer recurrence and/or following surgery (e.g., removal of the spleen). In some instances, a methylation score is utilized to determine the prognosis of a subject having colon cancer.

[0095] In some embodiments, a methylation score of from about 1.5 to about 4 is associated with a "good" prognosis. In some instances, a "good" prognosis refers to the likelihood that a subject will likely respond favorably to a drug or set of drugs, leading to a complete or partial remission of colon cancer or a decrease and/or a stop in the progression of colon cancer. In some instances, a "good" prognosis refers to the survival of a subject of from at least 1 month to at least 90 years. In some instances, a "good" prognosis refers to the survival of a subject in which the survival of the subject upon treatment is from at least 1 month to at least 90 years. In some instances, the survival of a subject further refers to an extended survival rate of a subject receiving a treatment course relative to a subject without receiving the same course of treatment. In some cases, a "good" prognosis refers to an extended survival time of a subject receiving a treatment course relative to a subject without receiving the same course of treatment.

[0096] In some instances, a methylation score of from about 1.5 to about 4 is indicative of a survival from at least 1 month to at least 90 years. In some instances, a methylation score of from about 1.5 to about 4 is indicative of a survival of at least 2 months, 4 months, 6 months, 8 months, 10 months, 1 year, 1.5 years, 2 years, 3 years, 4 years, 5 years, 10 years, 15 years, 20 years, 30 years, 50 years, or more.

[0097] In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival from at least 1 month to at least 90 years. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival of at least 2 months, 4 months, 6 months, 8 months, 10 months, 1 year, 1.5 years, 2 years, 3 years, 4 years, 5 years, 10 years, 15 years, 20 years, 30 years, 50 years, or more.

[0098] In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival from at least 1 month to at least 90 years. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival of at least 2 months, 4 months, 6 months, 8 months, 10 months, 1 year, 1.5 years, 2 years, 3 years, 4 years, 5 years, 10 years, 15 years, 20 years, 30 years, 50 years, or more.

[0099] In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival from at least 1 month to at least 90 years. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival of at least 2 months, 4 months, 6 months, 8 months, 10 months, 1 year, 1.5 years, 2 years, 3 years, 4 years, 5 years, 10 years, 15 years, 20 years, 30 years, 50 years, or more.

[0100] In some embodiments, a methylation score of from about 1.5 to about 4 is associated with a "good" prognosis in a subject having colon cancer. In some embodiments, a methylation score of from about 1.5 to about 4, from about 1.5 to about 3.5, from about 1.5 to about 3, from about 1.5 to about 2.5, or from about 1.5 to about 2 is associated with a "good" prognosis in a subject having colon cancer.

[0101] In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival from at least 1 month to at least 90 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival of at least 2 months, 4 months, 6 months, 8 months, 10 months, 1 year, 1.5 years, 2 years, 3 years, 4 years, 5 years, 10 years, 15 years, 20 years, 30 years, 50 years, or more in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 2 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 3 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 4 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 5 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 6 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 8 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 10 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 1 year in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 1.5 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 2 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 2.5 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 3 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 4 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 3 is indicative of a survival for at least 5 years in a subject having colon cancer.

[0102] In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival from at least 1 month to at least 90 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival of at least 2 months, 4 months, 6 months, 8 months, 10 months, 1 year, 1.5 years, 2 years, 3 years, 4 years, 5 years, 10 years, 15 years, 20 years, 30 years, 50 years, or more in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 2 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 3 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 4 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 5 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 6 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 8 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 10 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 1 year in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 1.5 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 2 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 2.5 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 3 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 4 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 5 years in a subject having colon cancer.

[0103] In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival from at least 1 month to at least 90 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival of at least 2 months, 4 months, 6 months, 8 months, 10 months, 1 year, 1.5 years, 2 years, 3 years, 4 years, 5 years, 10 years, 15 years, 20 years, 30 years, 50 years, or more in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 2 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 3 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 4 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 5 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 6 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 8 months in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 10 months in a subject having colon cancer.

[0104] In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 1 year in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 1.5 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 2 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 2.5 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 3 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 4 years in a subject having colon cancer. In some instances, a methylation score of from about 1.5 to about 2 is indicative of a survival for at least 5 years in a subject having colon cancer.

[0105] In some embodiments, a methylation score of less than about 1.5 is associated with a “poor” prognosis. In some instances, a “poor” prognosis refers to the likelihood that a subject will likely respond unfavorably to a drug or set of drugs, leading to a progression of colon cancer (e.g., progression to metastatic colon cancer) and/or to refractory of one or more therapeutic agents. In some instances, a “poor” prognosis refers to the likelihood that a subject will not respond to a drug or set of drugs, leading to a progression of colon cancer. In some instances, a “poor” prognosis refers to the survival of a subject of from less than 5 years to less than 1 month. In some instances, a “poor” prognosis refers to the survival of a subject in which the survival of the subject upon treatment is from less than 5 years to less than 1 month. In some instances, a “poor” prognosis further refers to the likelihood that a subject will develop a refractory colon cancer toward one or more drugs.

[0106] In some instances, a methylation score of less than 1.5 is indicative of a survival of from less than 5 years to less than 1 month. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 5 years, 4 years, 3 years, 2 years, 1.5 years, 1 year, 10 months, 8 months, 6 months, 4 months, or 2 months.

[0107] In some embodiments, a methylation score of less than about 1.5 is associated with a “poor” prognosis in a subject having colon cancer. In some embodiments, a methylation score of less than about 1.5 is associated with a “poor” prognosis in a subject having colon cancer.

[0108] In some instances, a methylation score of less than 1.5 is indicative of a survival of from less than 5 years to less than 1 month in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 5 years, 4 years, 3 years, 2 years, 1.5 years, 1 year, 10 months, 8 months, 6 months, 4 months, or 2 months in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 5 years in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 4 years in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 3 years in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 2.5 years in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 2 years in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 1.5 years in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 1 year in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 6 months in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 5 months in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival f of less than 4 months in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 3 months in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 2 months in a subject having colon cancer. In some instances, a methylation score of less than 1.5 is indicative of a survival of less than 1 month in a subject having colon cancer.

[0109] In some instances, one or more samples are obtained from a subject during the course of a treatment to monitor the progression of colon cancer in the subject. In some instances, the subject initially has a methylation score of from about 1.5 to about 3 and progressively during each subsequent testing has a lower methylation score. For example, a subject initially has a methylation score of 3 and during subsequent testings, has methylation scores of 2.5, 2, 1.5, or 1. In such cases, the subject is further tested to determine the progression of colon cancer (e.g., whether colon cancer has progressed into a metastatic state or into a refractory state) and a treatment course is optionally altered based on the changes in prognosis.

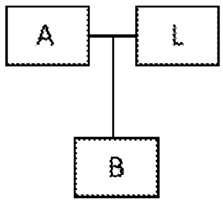
[0110] In some embodiments, the methylation score is calculated based on model for a survival analysis. In some instances, a survival analysis is a statistic analysis for analyzing the expected duration of time until one or more events of interest happen. In some instances, survival analysis comprises Cox proportional hazards (PH) regression analysis, log-rank test or a product limit estimator. In some instances, the methylation score is calculated based on Cox proportional hazards (PH) regression analysis, log-rank test or product limit estimator. In some instances, the methylation score is calculated based on Cox proportional hazards (PH) regression analysis. In some embodiments, the methylation score is further calculated based on a log-rank test. In some instances, the log-rank test is a hypothesis test to compare the survival distribution of two samples (e.g., a training set and a validation set). In some instances, the log-rank test is also referred to as a Mantel-Cox test or a time-stratified Cochran-Mantel-Haenszel test. In some instances, the methylation score is additionally calculated based on a product limit estimator. A product limit estimator (also known as Kaplan-Meier estimator) is a non-parametric statistic used to estimate the survival function from lifetime data. In some embodiments, the methylation score is initially calculated based on Cox proportional hazards (PH) regression analysis and then reprocessed with a log-rank test.

Control

[0111] In some embodiments, a control is a methylation value, methylation level, or methylation profile of a sample. In some instances, the control comprises a set of methylation profiles, wherein each said methylation profile is generated from a biological sample obtained from a known cancer type. In some cases, the known cancer type is colon cancer. In some cases, the known cancer type is a relapsed or refractory colon cancer. In other cases, the known cancer type is a metastatic colon cancer. In some cases, the known cancer type is adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma.

Probes

[0112] In some embodiments, one or more probes described above comprise a structure of Formula I:



Formula I

wherein:

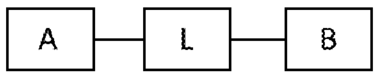
A is a first target-binding region;

B is a second target-binding region; and

L is a linker region;

wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 30 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15; B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 12 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15; and wherein L is attached to A; and B is attached to either A or L.

[0113] In some instances, L is attached to A and B is attached to L. In some cases, A, B, and L are attached as illustrated in Formula Ia:



Formula Ia.

[0114] In some embodiments, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 35 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 40 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 45 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID

NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 50 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 55 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 60 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 65 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 70 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 80 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15. In some cases, A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 90 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15.

[0115] In some embodiments, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 14 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 15 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 18 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 20 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 22 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 25 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 28 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from

SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 30 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 35 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 40 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 45 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 50 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 55 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 60 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 65 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 70 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 80 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15. In some cases, B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 90 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15.

[0116] In some instances, a probe described above is used in a next generation sequencing reaction to generate a CpG methylation data. In some instances, the probe is used in a solution-based next generation sequencing reaction to generate a CpG methylation data. In some instances, the next generation sequencing reaction comprises 454 Life Sciences platform (Roche, Branford, CT); lllumina's Genome Analyzer, GoldenGate Methylation Assay, or Infinium Methylation Assays, i.e., Infinium HumanMethylation 27K BeadArray or VeraCode GoldenGate methylation array (Illumina, San Diego, CA); QX200™ Droplet

Digital™ PCR System from Bio-Rad; DNA Sequencing by Ligation, SOLiD System (Applied Biosystems/Life Technologies); the Helicos True Single Molecule DNA sequencing technology; semiconductor sequencing (Ion Torrent; Personal Genome Machine); DNA nanoball sequencing; sequencing using technology from Dover Systems (Polonator), and technologies that do not require amplification or otherwise transform native DNA prior to sequencing (e.g., Pacific Biosciences and Helicos), such as nanopore-based strategies (e.g., Oxford Nanopore, Genia Technologies, and Nabsys). In some instances, the solution-based next generation sequencing reaction is a droplet digital PCR sequencing method.

[0117] In some instances, each probe correlates to a CpG site. In some instances, each probe correlates to a biomarker (e.g., CpG site) as illustrated in Table 5.

[0118] In some instances, L is between 10 and 60, 15 and 55, 20 and 50, 25 and 45, and 30 and 40 nucleotides in length. In some instances, L is about 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 nucleotides in length.

[0119] In some instances, L further comprises an adaptor region. In some instances, the adaptor region comprises a sequence used to identify each probe. In some instances as illustrated in Table 5, the adaptor region in each illustrative sequence is reflected by a series of N, in which each N is A, T, G, or C.

[0120] In some embodiments, a probe described herein comprises at least 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 50% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 60% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 70% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 80% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 85% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 90% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 91% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 92% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 93% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 94% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 95% sequence identity to a sequence selected from SEQ ID NOs: 1-

15. In some instances, the probe comprises at least 96% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 97% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 98% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises at least 99% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe comprises 100% sequence identity to a sequence selected from SEQ ID NOs: 1-15. In some instances, the probe consists of a sequence selected from SEQ ID NOs: 1-15.

[0121] In some cases, a probe described above is utilized in a digital PCR sequencing method. In some cases, the probe is utilized in a droplet digital PCR (ddPCR) sequencing method.

Detection Methods

[0122] In some embodiments, a number of methods are utilized to measure, detect, determine, identify, and characterize the methylation status/level of a gene or a biomarker (e.g., CpG island-containing region/fragment) in identifying a subject as having colon cancer, determining the colon cancer subtype, the prognosis of a subject having colon cancer, and the progression or regression of colon cancer in subject in the presence of a therapeutic agent.

[0123] In some instances, the methylation profile is generated from a biological sample isolated from an individual. In some embodiments, the biological sample is a biopsy. In some instances, the biological sample is a tissue sample. In some instances, the biological sample is a tissue biopsy sample. In some instances, the biological sample is a blood sample. In other instances, the biological sample is a cell-free biological sample. In other instances, the biological sample is a circulating tumor DNA sample. In one embodiment, the biological sample is a cell free biological sample containing circulating tumor DNA.

[0124] In some embodiments, a biomarker (or an epigenetic marker) is obtained from a liquid sample. In some embodiments, the liquid sample comprises blood and other liquid samples of biological origin (including, but not limited to, peripheral blood, sera, plasma, ascites, urine, cerebrospinal fluid (CSF), sputum, saliva, bone marrow, synovial fluid, aqueous humor, amniotic fluid, cerumen, breast milk, broncheoalveolar lavage fluid, semen, prostatic fluid, cowper's fluid or pre-ejaculatory fluid, female ejaculate, sweat, tears, cyst fluid, pleural and peritoneal fluid, pericardial fluid, ascites, lymph, chyme, chyle, bile, interstitial fluid, menses, pus, sebum, vomit, vaginal secretions/flushing, synovial fluid, mucosal secretion, stool water, pancreatic juice, lavage fluids from sinus cavities, bronchopulmonary aspirates, blastocyl cavity fluid, or umbilical cord blood. In some embodiments, the biological fluid is blood, a blood derivative or a blood fraction, e.g., serum or plasma. In a specific embodiment, a sample comprises a blood sample. In another embodiment, a serum sample is used. In another embodiment, a sample comprises urine. In some embodiments, the liquid sample also encompasses a sample that has been manipulated in any way after their procurement, such as by centrifugation, filtration, precipitation, dialysis, chromatography, treatment with reagents, washed, or enriched for certain cell populations.

[0125] In some embodiments, a biomarker (or an epigenetic marker) is obtained from a tissue sample. In some instances, a tissue corresponds to any cell(s). Different types of tissue correspond to different types of cells (e.g., liver, lung, blood, connective tissue, and the like), but also healthy cells vs. tumor cells or to tumor cells at various stages of neoplasia, or to displaced malignant tumor cells. In some embodiments, a tissue sample further encompasses a clinical sample, and also includes cells in culture, cell supernatants, organs, and the like. Samples also comprise fresh-frozen and/or formalin- fixed, paraffin-embedded tissue blocks, such as blocks prepared from clinical or pathological biopsies, prepared for pathological analysis or study by immunohistochemistry.

[0126] In some embodiments, a biomarker (or an epigenetic marker) is methylated or unmethylated in a normal sample (e.g., normal or control tissue without disease, or normal or control body fluid, stool, blood, serum, amniotic fluid), most importantly in healthy stool, blood, serum, amniotic fluid or other body fluid. In other embodiments, a biomarker (or an epigenetic marker) is hypomethylated or hypermethylated in a sample from a patient having or at risk of a disease (e.g., one or more indications described herein); for example, at a decreased or increased (respectively) methylation frequency of at least about 50%, at least about 60%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or about 100% in comparison to a normal sample. In one embodiment, a sample is also hypomethylated or hypermethylated in comparison to a previously obtained sample analysis of the same patient having or at risk of a disease (e.g., one or more indications described herein), particularly to compare progression of a disease.

[0127] In some embodiments, a methylome comprises a set of epigenetic markers or biomarkers, such as a biomarker described above. In some instances, a methylome that corresponds to the methylome of a tumor of an organism (e.g., a human) is classified as a tumor methylome. In some cases, a tumor methylome is determined using tumor tissue or cell-free (or protein-free) tumor DNA in a biological sample. Other examples of methylomes of interest include the methylomes of organs that contribute DNA into a bodily fluid (e.g. methylomes of tissue such as brain, breast, lung, the prostate, and the kidneys, plasma, etc.).

[0128] In some embodiments, a plasma methylome is the methylome determined from the plasma or serum of an animal (e.g., a human). In some instances, the plasma methylome is an example of a cell-free or protein-free methylome since plasma and serum include cell-free DNA. The plasma methylome is also an example of a mixed methylome since it is a mixture of tumor and other methylomes of interest. In some instances, the urine methylome is determined from the urine sample of a subject. In some cases, a cellular methylome corresponds to the methylome determined from cells (e.g., blood cells) of the patient. The methylome of the blood cells is called the blood cell methylome (or blood methylome).

[0129] In some embodiments, DNA (e.g., genomic DNA such as extracted genomic DNA or treated genomic DNA) is isolated by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample is disrupted and lysed by enzymatic, chemical or mechanical means. In some cases, the DNA solution is then cleared of proteins and other contaminants e.g. by digestion with proteinase K. The DNA is then recovered from the solution. In such cases, this is carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. In some instances, the choice of method is affected by several factors including time, expense and required quantity of DNA.

[0130] Wherein the sample DNA is not enclosed in a membrane (e.g. circulating DNA from a cell free sample such as blood or urine) methods standard in the art for the isolation and/or purification of DNA are optionally employed (See, for example, Bettegowda et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Sci. Transl. Med, 6(224): ra24. 2014). Such methods include the use of a protein degenerating reagent e.g. chaotropic salt e.g. guanidine hydrochloride or urea; or a detergent e.g. sodium dodecyl sulphate (SDS), cyanogen bromide. Alternative methods include but are not limited to ethanol precipitation or propanol precipitation, vacuum concentration amongst others by means of a centrifuge. In some cases, the person skilled in the art also make use of devices such as filter devices e.g. ultrafiltration, silica surfaces or membranes, magnetic particles, polystyrol particles, polystyrol surfaces, positively charged surfaces, and positively charged membranes, charged membranes, charged surfaces, charged switch membranes, charged switched surfaces.

[0131] In some instances, once the nucleic acids have been extracted, methylation analysis is carried out by any means known in the art. A variety of methylation analysis procedures are known in the art and may be used to practice the methods disclosed herein. These assays allow for determination of the methylation state of one or a plurality of CpG sites within a tissue sample. In addition, these methods may be used for absolute or relative quantification of methylated nucleic acids. Such methylation assays involve, among other techniques, two major steps. The first step is a methylation specific reaction or separation, such as (i) bisulfite treatment, (ii) methylation specific binding, or (iii) methylation specific restriction enzymes. The second major step involves (i) amplification and detection, or (ii) direct detection, by a variety of methods such as (a) PCR (sequence-specific amplification) such as Taqman(R), (b) DNA sequencing of untreated and bisulfite-treated DNA, (c) sequencing by ligation of dye-modified probes (including cyclic ligation and cleavage), (d) pyrosequencing, (e) single-molecule sequencing, (f) mass spectroscopy, or (g) Southern blot analysis.

[0132] Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA may be used, e.g., the method described by Sadri and Hornsby (1996, Nucl. Acids Res. 24:5058- 5059), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong and Laird, 1997, Nucleic Acids Res. 25:2532- 2534). COBRA analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA. 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. (Frommer et al, 1992, Proc. Nat. Acad. Sci. USA, 89, 1827-1831). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the CpG sites 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 micro-dissected 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 gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfo nation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

[0133] In an embodiment, the methylation profile of selected CpG sites is determined using methylation-Specific PCR (MSP). MSP 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, 1996, Proc. Nat. Acad. Sci. USA, 93, 9821-9826; U.S. Pat. Nos. 5,786,146, 6,017,704, 6,200,756, 6,265,171 (Herman and Baylin); U.S. Pat. Pub. No. 2010/0144836 (Van Engeland et al)). Briefly, DNA is modified by a deaminating agent such as sodium bisulfite to convert unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers specific for methylated versus unmethylated DNA. In some instances, typical reagents (e.g., as might be found in a typical MSP- based kit) for MSP analysis include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation- altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes. The ColoSure™ test is a commercially available test for colon cancer based on the MSP technology and measurement of methylation of the vimentin gene (Itzkowitz et al, 2007, Clin Gastroenterol. Hepatol. 5(1), 111-117). Alternatively, one may use quantitative multiplexed methylation specific PCR (QM-PCR), as described by Fackler et al. Fackler et al, 2004, Cancer Res. 64(13) 4442-4452; or Fackler et al, 2006, Clin. Cancer Res. 12(11 Pt 1) 3306-3310.

[0134] In an embodiment, the methylation profile of selected CpG sites is determined using MethyLight and/or Heavy Methyl Methods. The MethyLight and Heavy Methyl assays are a high-throughput quantitative methylation assay that utilizes fluorescence- based real-time PCR (Taq Man(R)) technology that requires no further manipulations after the PCR step (Eads, C.A. et al, 2000, Nucleic Acid Res. 28, e 32; Cottrell et al, 2007, J. Urology 177, 1753, U.S. Pat. Nos. 6,331,393 (Laird et al)). 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 either in an "unbiased" (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. In some cases, sequence discrimination occurs either at the level of the amplification process or at the level of the fluorescence detection process, or both. In some cases, the MethyLight assay is used as a quantitative test for methylation patterns in the genomic DNA sample, 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 of the biased PCR pool with either control oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites. Typical reagents (e.g., as might be found in a typical MethyLight- based kit) for MethyLight analysis include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan(R) probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase. In some instances, the MethyLight technology include epi proColon assay and mSEPT9 assay (Epigenomics, Berlin, Germany).

[0135] Quantitative MethyLight uses bisulfite to convert genomic DNA and the methylated sites are amplified using PCR with methylation independent primers. Detection probes specific for the methylated and unmethylated sites with two different fluorophores provides simultaneous quantitative measurement of the methylation. The Heavy Methyl technique begins with bisulfate conversion of DNA. Next specific blockers prevent the amplification of unmethylated DNA. Methylated genomic DNA does not bind the blockers and their sequences will be amplified. The amplified sequences are detected with a methylation specific probe. (Cottrell et al, 2004, Nuc. Acids Res. 32:e10, the contents of which is hereby incorporated by reference in its entirety).

[0136] 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 and Jones, 1997, Nucleic Acids Res. 25, 2529-2531). 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(s) of interest. In some cases, small amounts of DNA are analyzed (e.g., micro-dissected pathology sections), and the method avoids utilization of restriction enzymes for determining the methylation status at CpG sites. Typical reagents (e.g., as is found in a typical Ms-SNuPE-based kit) for Ms- SNuPE analysis include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

[0137] In another embodiment, the methylation status of selected CpG sites is determined using differential Binding-based Methylation Detection Methods. For identification of differentially methylated regions, one approach is to capture methylated DNA. This approach uses a protein, in which the methyl binding domain of MBD2 is fused to the Fc fragment of an antibody (MBD-FC) (Gebhard et al, 2006, Cancer Res. 66:6118-6128; and PCT Pub. No. WO 2006/056480 A2 (Relhi)). This fusion protein has several advantages over conventional methylation specific antibodies. The MBD FC has a higher affinity to methylated DNA and it binds double stranded DNA. Most importantly the two proteins differ in the way they bind DNA. Methylation specific antibodies bind DNA stochastically, which means that only a binary answer can be obtained. The methyl binding domain of MBD-FC, on the other hand, binds DNA molecules regardless of their methylation status. The strength of this protein - DNA interaction is defined by the level of DNA methylation. After binding genomic DNA, eluate solutions of increasing salt concentrations can be used to fractionate non- methylated and methylated DNA allowing for a more controlled separation (Gebhard et al, 2006, Nucleic Acids Res. 34: e82). Consequently this method, called Methyl-CpG immunoprecipitation (MCIP), not only enriches, but also fractionates genomic DNA according to methylation level, which is particularly helpful when the unmethylated DNA fraction should be investigated as well.

[0138] In an alternative embodiment, a 5 -methyl cytidine antibody to bind and precipitate methylated DNA. Antibodies are available from Abeam (Cambridge, MA), Diagenode (Sparta, NJ) or Eurogentec (c/o AnaSpec, Fremont, CA). Once the methylated fragments have been separated they may be sequenced using microarray based techniques such as methylated CpG-island recovery assay (MIRA) or methylated DNA immunoprecipitation (MeDIP) (Pelizzola et al, 2008, Genome Res. 18, 1652-1659; O'Geen et al, 2006, BioTechniques 41(5), 577-580, Weber et al, 2005, Nat. Genet. 37, 853-862; Horak and Snyder, 2002, Methods Enzymol, 350, 469-83; Lieb, 2003, Methods Mol Biol, 224, 99-109). Another technique is methyl-CpG binding domain column/segregation of partly melted molecules (MBD/SPM, Shiraishi et al, 1999, Proc. Natl. Acad. Sci. USA 96(6):2913-2918).

[0139] In some embodiments, methods for detecting methylation include randomly shearing or randomly fragmenting the genomic DNA, cutting the DNA with a methylation-dependent or methylation-sensitive restriction enzyme and subsequently selectively identifying and/or analyzing the cut or uncut DNA. Selective identification can include, for example, separating cut and uncut DNA (e.g., by size) and quantifying a sequence of interest that was cut or, alternatively, that was not cut. See, e.g., U.S. Patent No. 7,186,512. Alternatively, the method can encompass amplifying intact DNA after restriction enzyme digestion, thereby only amplifying DNA that was not cleaved by the restriction enzyme in the area amplified. See, e.g., U.S. Patents No. 7,910,296; No. 7,901,880; and No. 7,459,274. In some embodiments, amplification can be performed using primers that are gene specific.

[0140] For example, there are methyl-sensitive enzymes that preferentially or substantially cleave or digest at their DNA recognition sequence if it is non-methylated. Thus, an unmethylated DNA sample is cut into smaller fragments than a methylated DNA sample. Similarly, a hypermethylated DNA sample is not cleaved. In contrast, there are methyl-sensitive enzymes that cleave at their DNA recognition sequence only if it is methylated. Methyl- sensitive enzymes that digest unmethylated DNA suitable for use in methods of the technology include, but are not limited to, Hpall, Hhal, Maell, BstUI and Acil. In some instances, an enzyme that is used is Hpall that cuts only the unmethylated sequence CCGG. In other instances, another enzyme that is used is Hhal that cuts only the unmethylated sequence GCGC. Both enzymes are available from New England BioLabs(R), Inc. Combinations of two or more methyl-sensitive enzymes that digest only unmethylated DNA are also used. Suitable enzymes that digest only methylated DNA include, but are not limited to, Dpnl, which only cuts at fully methylated 5'-GATC sequences, and McrBC, an endonuclease, which cuts DNA containing modified cytosines (5-methylcytosine or 5-hydroxymethylcytosine or N4- methylcytosine) and cuts at recognition site 5'... PumC(N4o-3ooo) PumC... 3' (New England BioLabs, Inc., Beverly, MA). Cleavage methods and procedures for selected restriction enzymes for cutting DNA at specific sites are well known to the skilled artisan. For example, many suppliers of restriction enzymes provide information on conditions and types of DNA sequences cut by specific restriction enzymes, including New England BioLabs, Pro-Mega Biochems, Boehringer-Mannheim, and the like. Sambrook et al. (See Sambrook et al. Molecular Biology: A Laboratory Approach, Cold Spring Harbor, N.Y. 1989) provide a general description of methods for using restriction enzymes and other enzymes.

[0141] In some instances, a methylation-dependent restriction enzyme is a restriction enzyme that cleaves or digests DNA at or in proximity to a methylated recognition sequence, but does not cleave DNA at or near the same sequence when the recognition sequence is not methylated. Methylation-dependent restriction enzymes include those that cut at a methylated recognition sequence (e.g., Dpnl) and enzymes that cut at a sequence near but not at the recognition sequence (e.g., McrBC). For example, McrBC's recognition sequence is 5' RmC (N40-3000) RmC 3 ' where "R" is a purine and "mC" is a methylated cytosine and "N40-3000" indicates the distance between the two RmC half sites for which a restriction event has been observed. McrBC generally cuts close to one half-site or the other, but cleavage positions are typically distributed over several base pairs, approximately 30 base pairs from the methylated base. McrBC sometimes cuts 3' of both half sites, sometimes 5' of both half sites, and sometimes between the two sites. Exemplary methylation-dependent restriction enzymes include, e.g., McrBC, McrA, MrrA, Bisl, Glal and Dpnl. One of skill in the art will appreciate that any methylation-dependent restriction enzyme, including homologs and orthologs of the restriction enzymes described herein, is also suitable for use with one or more methods described herein.

[0142] In some cases, a methylation-sensitive restriction enzyme is a restriction enzyme that cleaves DNA at or in proximity to an unmethylated recognition sequence but does not cleave at or in proximity to the same sequence when the recognition sequence is methylated. Exemplary methylation-sensitive restriction enzymes are described in, e.g., McClelland et al, 22(17) NUCLEIC ACIDS RES. 3640-59 (1994). Suitable methylation-sensitive restriction enzymes that do not cleave DNA at or near their recognition sequence when a cytosine within the recognition sequence is methylated at position C5 include, e.g., Aat II, Aci I, Acd I, Age I, Alu I, Asc I, Ase I, AsiS I, Bbe I, BsaA I, BsaH I, BsiE I, BsiW I, BsrF I, BssH II, BssK I, BstB I, BstN I, BstU I, Cla I, Eae I, Eag I, Fau I, Fse I, Hha I, HinPl I, HinC II, Hpa II, Hpy99 I, HpyCH4 IV, Kas I, Mbo I, Mlu I, MapAl I, Msp I, Nae I, Nar I, Not I, Pml I, Pst I, Pvu I, Rsr II, Sac II, Sap I, Sau3A I, Sfl I, Sfo I, SgrA I, Sma I, SnaB I, Tsc I, Xma I, and Zra I. Suitable methylation-sensitive restriction enzymes that do not cleave DNA at or near their recognition sequence when an adenosine within the recognition sequence is methylated at position N6 include, e.g., Mbo I. One of skill in the art will appreciate that any methylation-sensitive restriction enzyme, including homologs and orthologs of the restriction enzymes described herein, is also suitable for use with one or more of the methods described herein. One of skill in the art will further appreciate that a methylation-sensitive restriction enzyme that fails to cut in the presence of methylation of a cytosine at or near its recognition sequence may be insensitive to the presence of methylation of an adenosine at or near its recognition sequence. Likewise, a methylation-sensitive restriction enzyme that fails to cut in the presence of methylation of an adenosine at or near its recognition sequence may be insensitive to the presence of methylation of a cytosine at or near its recognition sequence. For example, Sau3AI is sensitive (i.e., fails to cut) to the presence of a methylated cytosine at or near its recognition sequence, but is insensitive (i.e., cuts) to the presence of a methylated adenosine at or near its recognition sequence. One of skill in the art will also appreciate that some methylation-sensitive restriction enzymes are blocked by methylation of bases on one or both strands of DNA encompassing of their recognition sequence, while other methylation-sensitive restriction enzymes are blocked only by methylation on both strands, but can cut if a recognition site is hemi-methylated.

[0143] In alternative embodiments, adaptors are optionally added to the ends of the randomly fragmented DNA, the DNA is then digested with a methylation-dependent or methylation-sensitive restriction enzyme, and intact DNA is subsequently amplified using primers that hybridize to the adaptor sequences. In this case, a second step is performed to determine the presence, absence or quantity of a particular gene in an amplified pool of DNA. In some embodiments, the DNA is amplified using real-time, quantitative PCR.

[0144] In other embodiments, the methods comprise quantifying the average methylation density in a target sequence within a population of genomic DNA. In some embodiments, the method comprises contacting genomic DNA with a methylation-dependent restriction enzyme or methylation-sensitive restriction enzyme under conditions that allow for at least some copies of potential restriction enzyme cleavage sites in the locus to remain uncleaved; quantifying intact copies of the locus; and comparing the quantity of amplified product to a control value representing the quantity of methylation of control DNA, thereby quantifying the average methylation density in the locus compared to the methylation density of the control DNA.

[0145] In some instances, the quantity of methylation of a locus of DNA is determined by providing a sample of genomic DNA comprising the locus, cleaving the DNA with a restriction enzyme that is either methylation-sensitive or methylation-dependent, and then quantifying the amount of intact DNA or quantifying the amount of cut DNA at the DNA locus of interest. The amount of intact or cut DNA will depend on the initial amount of genomic DNA containing the locus, the amount of methylation in the locus, and the number (i.e., the fraction) of nucleotides in the locus that are methylated in the genomic DNA. The amount of methylation in a DNA locus can be determined by comparing the quantity of intact

DNA or cut DNA to a control value representing the quantity of intact DNA or cut DNA in a similarly-treated DNA sample. The control value can represent a known or predicted number of methylated nucleotides. Alternatively, the control value can represent the quantity of intact or cut DNA from the same locus in another (e.g., normal, non-diseased) cell or a second locus.

[0146] By using at least one methylation-sensitive or methylation-dependent restriction enzyme under conditions that allow for at least some copies of potential restriction enzyme cleavage sites in the locus to remain uncleaved and subsequently quantifying the remaining intact copies and comparing the quantity to a control, average methylation density of a locus can be determined. If the methylation-sensitive restriction enzyme is contacted to copies of a DNA locus under conditions that allow for at least some copies of potential restriction enzyme cleavage sites in the locus to remain uncleaved, then the remaining intact DNA will be directly proportional to the methylation density, and thus may be compared to a control to determine the relative methylation density of the locus in the sample. Similarly, if a methylation-dependent restriction enzyme is contacted to copies of a DNA locus under conditions that allow for at least some copies of potential restriction enzyme cleavage sites in the locus to remain uncleaved, then the remaining intact DNA will be inversely proportional to the methylation density, and thus may be compared to a control to determine the relative methylation density of the locus in the sample. Such assays are disclosed in, e.g., U.S. Patent No. 7,910,296.

[0147] The methylated CpG island amplification (MCA) technique is a method that can be used to screen for altered methylation patterns in genomic DNA, and to isolate specific sequences associated with these changes (Toyota et al, 1999, Cancer Res. 59, 2307-2312, U.S. Pat. No. 7,700,324 (Issa et al)). Briefly, restriction enzymes with different sensitivities to cytosine methylation in their recognition sites are used to digest genomic DNAs from primary tumors, cell lines, and normal tissues prior to arbitrarily primed PCR amplification. Fragments that show differential methylation are cloned and sequenced after resolving the PCR products on high-resolution polyacrylamide gels. The cloned fragments are then used as probes for Southern analysis to confirm differential methylation of these regions. Typical reagents (e.g., as might be found in a typical MCA-based kit) for MCA analysis may include, but are not limited to: PCR primers for arbitrary priming Genomic DNA; PCR buffers and nucleotides, restriction enzymes and appropriate buffers; gene-hybridization oligos or probes; control hybridization oligos or probes.

[0148] Additional methylation detection methods include those methods described in, e.g., U.S. Patents No. 7,553,627; No. 6,331,393; U.S. Patent Serial No. 12/476,981; U.S. Patent Publication No. 2005/0069879; Rein, et al, 26(10) NUCLEIC ACIDS RES. 2255-64 (1998); and Olek et al, 17(3) NAT. GENET. 275-6 (1997).

[0149] In another embodiment, the methylation status of selected CpG sites is determined using Methylation- Sensitive High Resolution Melting (HRM). Recently, Wojdacz et al. reported methylation-sensitive high resolution melting as a technique to assess methylation. (Wojdacz and Dobrovic, 2007, Nuc. Acids Res. 35(6) e41; Wojdacz et al. 2008, Nat. Prot. 3(12) 1903-1908; Balic et al, 2009 J. Mol. Diagn. 11 102- 108; and US Pat. Pub. No. 2009/0155791 (Wojdacz et al)). A variety of commercially available real time PCR machines have HRM systems including the Roche LightCycler480, Corbett Research RotorGene6000, and the Applied Biosystems 7500. HRM may also be combined with other amplification techniques such as pyrosequencing as described by Candiloro et al. (Candiloro et al, 2011, Epigenetics 6(4) 500-507).

[0150] In another embodiment, the methylation status of selected CpG locus is determined using a primer extension assay, including an optimized PCR amplification reaction that produces amplified targets for analysis using mass spectrometry. The assay can also be done in multiplex. Mass spectrometry is a particularly effective method for the detection of polynucleotides associated with the differentially methylated regulatory elements. The presence of the polynucleotide sequence is verified by comparing the mass of the detected signal with the expected mass of the polynucleotide of interest. The relative signal strength, e.g., mass peak on a spectra, for a particular polynucleotide sequence indicates the relative population of a specific allele, thus enabling calculation of the allele ratio directly from the data. This method is described in detail in PCT Pub. No. WO 2005/012578A1 (Beaulieu et al), which is hereby incorporated by reference in its entirety. For methylation analysis, the assay can be adopted to detect bisulfite introduced methylation dependent C to T sequence changes. These methods are particularly useful for performing multiplexed amplification reactions and multiplexed primer extension reactions (e.g., multiplexed homogeneous primer mass extension (hME) assays) in a single well to further increase the throughput and reduce the cost per reaction for primer extension reactions.

[0151] Other methods for DNA methylation analysis include restriction landmark genomic scanning (RLGS, Costello et al, 2002, Meth. Mol Biol, 200, 53-70), methylation- sensitive-representational difference analysis (MS-RDA, Ushijima and Yamashita, 2009, Methods Mol Biol 507, 1 17-130). Comprehensive high-throughput arrays for relative methylation

(CHARM) techniques are described in WO 2009/021141 (Feinberg and Irizarry). The Roche(R) NimbleGen(R) microarrays including the Chromatin Immunoprecipitation-on- chip (ChlP-chip) or methylated DNA immunoprecipitation-on-chip (MeDIP-chip). These tools have been used for a variety of cancer applications including melanoma, liver cancer and lung cancer (Koga et al, 2009, Genome Res., 19, 1462-1470; Acevedo et al, 2008, Cancer Res., 68, 2641-2651; Rauch et al, 2008, Proc. Nat. Acad. Sci. USA, 105, 252-257). Others have reported bisulfate conversion, padlock probe hybridization, circularization, amplification and next generation or multiplexed sequencing for high throughput detection of methylation (Deng et al, 2009, Nat. Biotechnol 27, 353-360; Ball et al, 2009, Nat. Biotechnol 27, 361-368; U.S. Pat. No. 7,611,869 (Fan)). As an alternative to bisulfate oxidation, Bayeyt et al. have reported selective oxidants that oxidize 5-methylcytosine, without reacting with thymidine, which are followed by PCR or pyro sequencing (WO 2009/049916 (Bayeyt et al).

[0152] In some instances, quantitative amplification methods (e.g., quantitative PCR or quantitative linear amplification) are used to quantify the amount of intact DNA within a locus flanked by amplification primers following restriction digestion. Methods of quantitative amplification are disclosed in, e.g., U.S. Patents No. 6, 180,349; No. 6,033,854; and No. 5,972,602, as well as in, e.g., DeGraves, et al, 34(1) BIOTECHNIQUES 106-15 (2003); Deiman B, et al., 20(2) MOL. BIOTECHNOL. 163-79 (2002); and Gibson et al, 6 GENOME RESEARCH 995-1001 (1996).

[0153] Following reaction or separation of nucleic acid in a methylation specific manner, the nucleic acid in some cases are subjected to sequence-based analysis. For example, once it is determined that one particular genomic sequence from a sample is hypermethylated or hypomethylated compared to its counterpart, the amount of this genomic sequence can be determined. Subsequently, this amount can be compared to a standard control value and used to determine the present of colon cancer in the sample. In many instances, it is desirable to amplify a nucleic acid sequence using any of several nucleic acid amplification procedures which are well known in the art. Specifically, nucleic acid amplification is the chemical or enzymatic synthesis of nucleic acid copies which contain a sequence that is complementary to a nucleic acid sequence being amplified (template). The methods and kits may use any nucleic acid amplification or detection methods known to one skilled in the art, such as those described in U.S. Pat. Nos. 5,525,462 (Takarada et al); 6,114,117 (Hepp et al); 6,127,120 (Graham et al); 6,344,317 (Urnovitz); 6,448,001 (Oku); 6,528,632 (Catanzariti et al); and PCT Pub. No. WO 2005/111209 (Nakajima et al).

[0154] In some embodiments, the nucleic acids are amplified by PCR amplification using methodologies known to one skilled in the art. One skilled in the art will recognize, however, that amplification can be accomplished by any known method, such as ligase chain reaction (LCR), Q -replicas amplification, rolling circle amplification, transcription amplification, self-sustained sequence replication, nucleic acid sequence-based amplification (NASBA), each of which provides sufficient amplification. Branched-DNA technology is also optionally used to qualitatively demonstrate the presence of a sequence of the technology, which represents a particular methylation pattern, or to quantitatively determine the amount of this particular genomic sequence in a sample. Nolte reviews branched-DNA signal amplification for direct quantitation of nucleic acid sequences in clinical samples (Nolte, 1998, Adv. Clin. Chem. 33:201-235).

[0155] The PCR process is well known in the art and include, for example, reverse transcription PCR, ligation mediated PCR, digital PCR (dPCR), or droplet digital PCR (ddPCR). For a review of PCR methods and protocols, see, e.g., Innis et al, eds., PCR Protocols, A Guide to Methods and Application, Academic Press, Inc., San Diego, Calif. 1990; U.S. Pat. No. 4,683,202 (Mullis). PCR reagents and protocols are also available from commercial vendors, such as Roche Molecular Systems. In some instances, PCR is carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing region, a primer annealing region, and an extension reaction region automatically. Machines specifically adapted for this purpose are commercially available.

[0156] In some embodiments, amplified sequences are also measured using invasive cleavage reactions such as the Invader(R) technology (Zou et al, 2010, Association of Clinical Chemistry (AACC) poster presentation on July 28, 2010, “Sensitive Quantification of Methylated Markers with a Novel Methylation Specific Technology; and U.S. Pat. No. 7,011,944 (Prudent et al)).

[0157] Suitable next generation sequencing technologies are widely available. Examples include the 454 Life Sciences platform (Roche, Branford, CT) (Margulies et al. 2005 Nature, 437, 376-380); lllumina’s Genome Analyzer, GoldenGate Methylation Assay, or Infinium Methylation Assays, i.e., Infinium HumanMethylation 27K BeadArray or VeraCode GoldenGate methylation array (Illumina, San Diego, CA; Bibkova et al, 2006, Genome Res. 16, 383-393; U.S. Pat. Nos. 6,306,597 and 7,598,035 (Macevicz); 7,232,656 (Balasubramanian et al.)); QX200™ Droplet Digital™ PCR System from Bio-Rad; or DNA Sequencing by Ligation, SOLiD System (Applied Biosystems/Life Technologies; U.S. Pat.

Nos. 6,797,470, 7,083,917, 7,166,434, 7,320,865, 7,332,285, 7,364,858, and 7,429,453 (Barany et al); the Helicos True Single Molecule DNA sequencing technology (Harris et al, 2008 Science, 320, 106-109; U.S. Pat. Nos. 7,037,687 and 7,645,596 (Williams et al); 7, 169,560 (Lapidus et al); 7,769,400 (Harris)), the single molecule, real-time (SMRT™) technology of Pacific Biosciences, and sequencing (Soni and Meller, 2007, Clin. Chem. 53, 1996-2001); semiconductor sequencing (Ion Torrent; Personal Genome Machine); DNA nanoball sequencing; sequencing using technology from Dover Systems (Polonator), and technologies that do not require amplification or otherwise transform native DNA prior to sequencing (e.g., Pacific Biosciences and Helicos), such as nanopore-based strategies (e.g., Oxford Nanopore, Genia Technologies, and Nabsys). These systems allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion. Each of these platforms allow sequencing of clonally expanded or non-amplified single molecules of nucleic acid fragments. Certain platforms involve, for example, (i) sequencing by ligation of dye- modified probes (including cyclic ligation and cleavage), (ii) pyrosequencing, and (iii) single-molecule sequencing.

[0158] Pyrosequencing is a nucleic acid sequencing method based on sequencing by synthesis, which relies on detection of a pyrophosphate released on nucleotide incorporation. Generally, sequencing by synthesis involves synthesizing, one nucleotide at a time, a DNA strand complimentary to the strand whose sequence is being sought. Study nucleic acids may be immobilized to a solid support, hybridized with a sequencing primer, incubated with DNA polymerase, ATP sulfurylase, luciferase, apyrase, adenosine 5' phosphsulfate and luciferin. Nucleotide solutions are sequentially added and removed. Correct incorporation of a nucleotide releases a pyrophosphate, which interacts with ATP sulfurylase and produces ATP in the presence of adenosine 5' phosphsulfate, fueling the luciferin reaction, which produces a chemiluminescent signal allowing sequence determination. Machines for pyrosequencing and methylation specific reagents are available from Qiagen, Inc. (Valencia, CA). See also Tost and Gut, 2007, Nat. Prot. 2 2265-2275. An example of a system that can be used by a person of ordinary skill based on pyrosequencing generally involves the following steps: ligating an adaptor nucleic acid to a study nucleic acid and hybridizing the study nucleic acid to a bead; amplifying a nucleotide sequence in the study nucleic acid in an emulsion; sorting beads using a picoliter multiwell solid support; and sequencing amplified nucleotide sequences by pyrosequencing methodology (e.g., Nakano et al, 2003, J. Biotech. 102, 117-124). Such a system can be used to exponentially amplify amplification products generated by a process described herein, e.g., by ligating a heterologous nucleic acid to the first amplification product generated by a process described herein.

CpG Methylation Data Analysis Methods

[0159] In certain embodiments, the methylation values measured for biomarkers of a biomarker panel are mathematically combined and the combined value is correlated to the underlying diagnostic question. In some instances, methylated biomarker values are combined by any appropriate state of the art mathematical method. Well-known mathematical methods for correlating a biomarker combination to a disease status employ methods like discriminant analysis (DA) (e.g., linear-, quadratic-, regularized-DA), Discriminant Functional Analysis (DFA), Kernel Methods (e.g., SVM), Multidimensional Scaling (MDS), Nonparametric Methods (e.g., k-Nearest-Neighbor Classifiers), PLS (Partial Least Squares), Tree-Based Methods (e.g., Logic Regression, CART, Random Forest Methods, Boosting/Bagging Methods), Generalized Linear Models (e.g., Logistic Regression), Principal Components based Methods (e.g., SIMCA), Generalized Additive Models, Fuzzy Logic based Methods, Neural Networks and Genetic Algorithms based Methods. The skilled artisan will have no problem in selecting an appropriate method to evaluate an epigenetic marker or biomarker combination described herein. In one embodiment, the method used in a correlating methylation status of an epigenetic marker or biomarker combination, e.g. to diagnose colon cancer or a colon cancer subtype, is selected from DA (e.g., Linear-, Quadratic-, Regularized Discriminant Analysis), DFA, Kernel Methods (e.g., SVM), MDS, Nonparametric Methods (e.g., k-Nearest-Neighbor Classifiers), PLS (Partial Least Squares), Tree-Based Methods (e.g., Logic Regression, CART, Random Forest Methods, Boosting Methods), or Generalized Linear Models (e.g., Logistic Regression), and Principal Components Analysis. Details relating to these statistical methods are found in the following references: Ruczinski et al., 12 J. OF COMPUTATIONAL AND GRAPHICAL STATISTICS 475-511 (2003); Friedman, J. H., 84 J. OF THE AMERICAN STATISTICAL ASSOCIATION 165-75 (1989); Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, The Elements of Statistical Learning, Springer Series in Statistics (2001); Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. Classification and regression trees, California: Wadsworth (1984); Breiman, L., 45 MACHINE LEARNING 5-32 (2001); Pepe, M. S., The Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford Statistical Science Series, 28 (2003); and Duda, R. O., Hart, P. E., Stork, D. O., Pattern Classification, Wiley Interscience, 2nd Edition (2001).

[0160] In one embodiment, the correlated results for each methylation panel are rated by their correlation to the disease or tumor type positive state, such as for example, by p-value test or t-value test or F-test. Rated (best first, i.e. low p- or t-value) biomarkers are then subsequently selected and added to the methylation panel until a certain diagnostic value is reached. Such methods include identification of methylation panels, or more broadly, genes that were differentially methylated among several classes using, for example, a random-variance t-test (Wright G. W. and Simon R, Bioinformatics 19:2448-2455,2003). Other methods include the step of specifying a significance level to be used for determining the epigenetic markers that will be included in the biomarker panel. Epigenetic markers that are differentially methylated between the classes at a univariate parametric significance level less than the specified threshold are included in the panel. It doesn't matter whether the specified significance level is small enough to exclude enough false discoveries. In some problems better prediction is achieved by being more liberal about the biomarker panels used as features. In some cases, the panels are biologically interpretable and clinically applicable, however, if fewer markers are included. Similar to cross-validation, biomarker selection is repeated for each training set created in the cross-validation process. That is for the purpose of providing an unbiased estimate of prediction error. The methylation panel for use with new patient sample data is the one resulting from application of the methylation selection and classifier of the "known" methylation information, or control methylation panel.

[0161] Models for utilizing methylation profile to predict the class of future samples can also be used. These models may be based on the Compound Covariate Predictor (Radmacher et al. Journal of Computational Biology 9:505-511, 2002), Diagonal Linear Discriminant Analysis (Dudoit et al. Journal of the American Statistical Association 97:77-87, 2002), Nearest Neighbor Classification (also Dudoit et al.), and Support Vector Machines with linear kernel (Ramaswamy et al. PNAS USA 98:15149-54, 2001). The models incorporated markers that were differentially methylated at a given significance level (e.g. 0.01, 0.05 or 0.1) as assessed by the random variance t-test (Wright G. W. and Simon R. Bioinformatics 19:2448-2455, 2003). The prediction error of each model using cross validation, preferably leave-one-out cross-validation (Simon et al. Journal of the National Cancer Institute 95:14-18, 2003can be estimated. For each leave-one-out cross-validation training set, the entire model building process is repeated, including the epigenetic marker selection process. In some instances, it is also evaluated in whether the cross-validated error rate estimate for a model is significantly less than one would expect from random prediction. In some cases, the class labels are randomly permuted and the entire leave-one-out cross-validation process is then repeated. The significance level is the proportion of the random permutations that gives a cross-validated error rate no greater than the cross-validated error rate obtained with the real methylation data.

[0162] Another classification method is the greedy-pairs method described by Bo and Jonassen (Genome Biology 3(4):research0017.1-0017.11, 2002). The greedy-pairs approach starts with ranking all markers based on their individual t-scores on the training set. This method attempts to select pairs of markers that work well together to discriminate the classes.

[0163] Furthermore, a binary tree classifier for utilizing methylation profile is optionally used to predict the class of future samples. The first node of the tree incorporated a binary classifier that distinguished two subsets of the total set of classes. The individual binary classifiers are based on the "Support Vector Machines" incorporating markers that were differentially expressed among markers at the significance level (e.g. 0.01, 0.05 or 0.1) as assessed by the random variance t-test (Wright G. W. and Simon R. Bioinformatics 19:2448-2455, 2003). Classifiers for all possible binary partitions are evaluated and the partition selected is that for which the cross-validated prediction error is minimum. The process is then repeated successively for the two subsets of classes determined by the previous binary split. The prediction error of the binary tree classifier can be estimated by cross-validating the entire tree building process. This overall cross-validation includes re-selection of the optimal partitions at each node and re-selection of the markers used for each cross-validated training set as described by Simon et al. (Simon et al. Journal of the National Cancer Institute 95:14-18, 2003). Several-fold cross validation in which a fraction of the samples is withheld, a binary tree developed on the remaining samples, and then class membership is predicted for the samples withheld. This is repeated several times, each time withholding a different percentage of the samples. The samples are randomly partitioned into fractional test sets (Simon R and Lam A. BRB-ArrayTools User Guide, version 3.2. Biometric Research Branch, National Cancer Institute).

[0164] Thus, in one embodiment, the correlated results for each marker b) are rated by their correct correlation to the disease, preferably by p-value test. It is also possible to include a step in that the markers are selected d) in order of their rating.

[0165] In additional embodiments, factors such as the value, level, feature, characteristic, property, etc. of a transcription rate, mRNA level, translation rate, protein level, biological activity, cellular characteristic or property, genotype, phenotype, etc. can be utilized in addition prior to, during, or after administering a therapy to a patient to enable further analysis of the patient's cancer status.

[0166] In some embodiments, a diagnostic test to correctly predict status is measured as the sensitivity of the assay, the specificity of the assay or the area under a receiver operated characteristic ("ROC") curve. In some instances, sensitivity is the percentage of true positives that are predicted by a test to be positive, while specificity is the percentage of true negatives that are predicted by a test to be negative. In some cases, an ROC curve provides the sensitivity of a test as a function of 1- specificity. The greater the area under the ROC curve, for example, the more accurate or powerful the predictive value of the test. Other useful measures of the utility of a test include positive predictive value and negative predictive value. Positive predictive value is the percentage of people who test positive that are actually positive. Negative predictive value is the percentage of people who test negative that are actually negative.

[0167] In some embodiments, one or more of the biomarkers disclosed herein show a statistical difference in different samples of at least $p<0.05$, $p<10^{-2}$, $p<10^{-3}$, $p<10^{-4}$ or $p<10^{-5}$. Diagnostic tests that use these biomarkers may show an ROC of at least 0.6, at least about 0.7, at least about 0.8, or at least about 0.9. In some instances, the biomarkers are differentially methylated in different subjects with or without colon cancer. In additional instances, the biomarkers for different subtypes of colon cancer are differentially methylated. In certain embodiments, the biomarkers are measured in a patient sample using the methods described herein and compared, for example, to predefined biomarker levels and are used to determine whether the patient has colon cancer, which colon cancer subtype does the patient have, and/or what is the prognosis of the patient having colon cancer. In other embodiments, the correlation of a combination of biomarkers in a patient sample is compared, for example, to a predefined set of biomarkers. In some embodiments, the measurement(s) is then compared with a relevant diagnostic amount(s), cut-off(s), or multivariate model scores that distinguish between the presence or absence of colon cancer, between colon cancer subtypes, and between a "good" or a "poor" prognosis. As is well understood in the art, by adjusting the particular diagnostic cut-off(s) used in an assay, one can increase sensitivity or specificity of the diagnostic assay depending on the preference of the diagnostician. In some embodiments, the particular diagnostic cut-off is determined, for example, by measuring the amount of biomarker hypermethylation or hypomethylation in a statistically significant number of samples from patients with or without colon cancer and from patients with different colon cancer subtypes, and drawing the cut-off to suit the desired levels of specificity and sensitivity.

Kits/Article of Manufacture

[0168] In some embodiments, provided herein include kits for detecting and/or characterizing the methylation profile of a biomarker described herein. In some instances, the kit comprises a plurality of primers or probes to detect or measure the methylation status/levels of one or more samples. Such kits comprise, in some instances, at least one polynucleotide that hybridizes to at least one of the methylation marker sequences described herein and at least one reagent for detection of gene methylation. Reagents for detection of methylation include, e.g., sodium bisulfate, polynucleotides designed to hybridize to sequence that is the product of a marker sequence if the marker sequence is not methylated (e.g., containing at least one C-U conversion), and/or a methylation- sensitive or methylation-dependent restriction enzyme. In some cases, the kits provide solid supports in the form of an assay apparatus that is adapted to use in the assay. In some instances, the kits further comprise detectable labels, optionally linked to a polynucleotide, e.g., a probe, in the kit.

[0169] In some embodiments, the kits comprise one or more (e.g., 1, 2, 3, 4, or more) different polynucleotides (e.g., primers and/or probes) capable of specifically amplifying at least a portion of a DNA region of a biomarker described herein. Optionally, one or more detectably-labeled polypeptides capable of hybridizing to the amplified portion are also included in the kit. In some embodiments, the kits comprise sufficient primers to amplify 2, 3, 4, 5, 6, 7, 8, 9, 10, or more different DNA regions or portions thereof, and optionally include detectably-labeled polynucleotides capable of hybridizing to each amplified DNA region or portion thereof. The kits further can comprise a methylation-dependent or methylation sensitive restriction enzyme and/or sodium bisulfite.

[0170] In some embodiments, the kits comprise sodium bisulfite, primers and adapters (e.g., oligonucleotides that can be ligated or otherwise linked to genomic fragments) for whole genome amplification, and polynucleotides (e.g., detectably-labeled polynucleotides) to quantify the presence of the converted methylated and or the converted unmethylated sequence of at least one cytosine from a DNA region of an epigenetic marker described herein.

[0171] In some embodiments, the kits comprise methylation sensing restriction enzymes (e.g., a methylation-dependent restriction enzyme and/or a methylation-sensitive restriction enzyme), primers and adapters for whole genome amplification, and polynucleotides to quantify the number of copies of at least a portion of a DNA region of an epigenetic marker described herein.

[0172] In some embodiments, the kits comprise a methylation binding moiety and one or more polynucleotides to quantify the number of copies of at least a portion of a DNA region of a marker described herein. A methylation binding moiety refers to a molecule (e.g., a polypeptide) that specifically binds to methyl-cytosine.

[0173] Examples include restriction enzymes or fragments thereof that lack DNA cutting activity but retain the ability to bind methylated DNA, antibodies that specifically bind to methylated DNA, etc.).

[0174] In some embodiments, the kit includes a packaging material. As used herein, the term "packaging material" can refer to a physical structure housing the components of the kit. In some instances, the packaging material maintains sterility of the kit components, and is made of material commonly used for such purposes (e.g., paper, corrugated fiber, glass, plastic, foil, ampules, etc.). Other materials useful in the performance of the assays are included in the kits, including test tubes, transfer pipettes, and the like. In some cases, the kits also include written instructions for the use of one or more of these reagents in any of the assays described herein.

[0175] In some embodiments, kits also include a buffering agent, a preservative, or a protein/nucleic acid stabilizing agent. In some cases, kits also include other components of a reaction mixture as described herein. For example, kits include one or more aliquots of thermostable DNA polymerase as described herein, and/or one or more aliquots of dNTPs. In some cases, kits also include control samples of known amounts of template DNA molecules harboring the individual alleles of a locus. In some embodiments, the kit includes a negative control sample, e.g., a sample that does not contain DNA molecules harboring the individual alleles of a locus. In some embodiments, the kit includes a positive control sample, e.g., a sample containing known amounts of one or more of the individual alleles of a locus.

Certain Terminology

[0176] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the claimed subject matter belongs. It is to be understood that the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of any subject matter claimed. In this application, the use of the singular includes the plural unless specifically stated otherwise. It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, use of the term “including” as well as other forms, such as “include”, “includes,” and “included,” is not limiting.

[0177] As used herein, ranges and amounts can be expressed as “about” a particular value or range. About also includes the exact amount. Hence “about 5 µL” means “about 5 µL” and also “5 µL.” Generally, the term “about” includes an amount that would be expected to be within experimental error.

[0178] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0179] As used herein, the terms “individual(s)”, “subject(s)” and “patient(s)” mean any mammal. In some embodiments, the mammal is a human. In some embodiments, the mammal is a non-human. None of the terms require or are limited to situations characterized by the supervision (e.g. constant or intermittent) of a health care worker (e.g. a doctor, a registered nurse, a nurse practitioner, a physician’s assistant, an orderly or a hospice worker).

[0180] A “site” corresponds to a single site, which in some cases is a single base position or a group of correlated base positions, e.g., a CpG site. A “locus” corresponds to a region that includes multiple sites. In some instances, a locus includes one site.

EXAMPLES

[0181] These examples are provided for illustrative purposes only and not to limit the scope of the claims provided herein.

Example 1. General Methodology

[0182] Tumor DNA extraction

[0183] Genomic DNA extraction from pieces of freshly frozen healthy or cancer tissues was performed with QIAamp DNA Mini Kit (Qiagen) according to manufacturer’s recommendations. DNA was extracted from roughly 0.5 mg of tissue. DNA was stored at -20^{O}C and analyzed within one week of preparation.

[0184] DNA extraction from FFPE samples

[0185] Genomic DNA from frozen FFPE samples was extracted using QIAamp DNA FFPE Tissue Kit with several modifications. DNA were stored at -20^{O}C for further analysis.

[0186] Bisulfite conversion of genomic DNA

[0187] 1µg of genomic DNA was converted to bis-DNA using EZ DNA Methylation-Lightning™ Kit (Zymo Research) according to the manufacturer’s protocol. Resulting bis-DNA had a size distribution of ~200-3000 bp, with a peak around ~500-1000 bp. The efficiency of bisulfite conversion was >99.8% as verified by deep-sequencing of bis-DNA and analyzing the ratio of C to T conversion of CH (non-CG) dinucleotides.

[0188] Determination of DNA methylation levels of the second validation cohort by deep sequencing of bis-DNA captured with molecular-inversion (padlock) probes

[0189] Padlock probes were designed to capture regions containing the CpG markers whose methylation levels significantly differed in any of the comparison between any cancer tissue and any normal tissue.

[0190] Probe design and synthesis

[0191] Padlock probes were designed using the ppDesigner software. The average length of the captured region was 100 bp, with the CpG marker located in the central portion of the captured region. Linker sequence between arms contained binding sequences for amplification primers separated by a variable stretch of Cs to produced probes of equal length. A 6-bp unique molecular identifier (UMI) sequence in probe design was incorporated to allow for the identification of unique individual molecular capture events and accurate scoring of DNA methylation levels. Probes were synthesized as separate oligonucleotides using standard commercial synthesis methods (ITD).

[0192] Bis-DNA capture

[0193] 50 ng of bisulfite-converted DNA was mixed with padlock probes in 20 μl reactions containing 1X Ampligase buffer (Epicentre). To anneal probes to DNA, 30 second denaturation at 95°C was followed by a slow cooling to 55°C. Hybridization was left to complete for 15 hrs at 55°C. To fill gaps between annealed arms, 5μl of the following mixture was added to each reaction: 2U of PfuTurbo polymerase (Agilent), 0.5U of Ampligase (Epicentre) and 250 pmol of each dNTP in 1X Ampligase buffer. 5 μl of exonuclease mix (20U of Exo I and 100U of ExoIII, both from Epicentre) was added and single-stranded DNA degradation was carried out at 37°C for 2 hours, followed by enzyme inactivation for 2 minutes at 94°C.

[0194] Circular products of site-specific capture were amplified by PCR with concomitant barcoding of separate samples. Amplification was carried out using primers specific to linker DNA within padlock probes, one of which contained specific 6bp barcodes. Both primers contained Illumina next-generation sequencing adaptor sequences. PCR was done as follows: 1X Phusion Flash Master Mix, 3 μl of captured DNA and 200nM primers, using the following cycle: 10s @ 98°C, 8X of (1s @ 98°C, 5s @ 58°C, 10s @ 72°C), 25X of (1s @ 98°C, 15s @ 72°C), 60s @ 72°C. PCR reactions were mixed and the resulting library was size selected to include effective captures (~230bp) and exclude "empty" captures (~150bp) using Agencourt AMPure XP beads (Beckman Coulter). Libraries were sequenced using MiSeq and HiSeq2500 systems (Illumina).

[0195] Sequencing data analysis

[0196] Mapping of sequencing reads was done using the software tool bisReadMapper with some modifications. First, UMI were extracted from each sequencing read and appended to read headers within FASTQ files using a custom script. Reads were on-the-fly converted as if all C were non-methylated and mapped to in-silico converted DNA strands of the human genome, also as if all C were non-methylated, using Bowtie2. Methylation frequencies were calculated for all CpG dinucleotides contained within the regions captured by padlock probes by dividing the numbers of unique reads carrying a C at the interrogated position by the total number of reads covering the interrogated position.

Example 2. Colon Cancer Diagnostic

[0197] Patient data was obtained from the Cancer Genome Atlas (TCGA). DNA methylation data were obtained from the TCGA analysis of about 450,000 sites generated using the Infinium 450K Methylation Array. Methylation profiles for colon cancer tissue and normal colon tissue were analyzed. Four clinical covariates were used, which includes:

[0198] Age (continuous);

[0199] Gender (categorical) with two levels: Female/ Male;

[0200] Race (categorical) with three levels: Asian, Black or African American and White. The category American Indian or Alaska Native was removed due to insufficient number of observations;

[0201] American Join Committee on Cancer (AJCC) stage - combined into a four-level covariate: Stage I, Stage II, Stage III and Stage IV.

[0202] The data were further modified by removal of missing value in any of the four clinical and demographic covariates to generate 404 colon cancer samples and 45 normal colon tissue samples for subsequent diagnostic analysis.

[0203] Six additional datasets were also obtained from TCGA which includes:

[0204] breast cancer tissue: 790

[0205] breast normal tissue: 97

[0206] liver cancer tissue: 377

[0207] liver normal tissue: 50

[0208] lung cancer tissue: 839

[0209] lung normal tissue: 74

[0210] and were used during the subsequent diagnostic analysis.

[0211] For each of the eight types of sample, each dataset (or sample set) were split into a training set and a test set with a 2:1 ratio. A pre-screening procedure was used first to remove excessive noise on the training data using the 'moderated t-statistics ' (Smyth, G. "Linear models and empirical bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology* **3**(1): 1-25 (2004)). For each set of comparison, one type of sample was compared against all other 7 types of samples. A list of markers with significantly difference in mean among all 8 sets of comparisons were retained for future analysis. The Benjamini-Hochberg procedure (Benjamini and Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B Methodological* p289-300, 1995) was used to control the FDR at significance level 0.05. For multinomial classification, least absolute shrinkage and selection operator (LASSO) was used under multinomial distribution. The tuning parameter was determined by the expected generalization error estimated from 10-fold cross-validation. Similar to survival analysis in Example 2, the random split scheme was repeated for 10 times to stabilize the variable selection procedure. A composite panel was constructed by keeping markers with a high selection probability and disregard those with low selection probability from the 10 sets of markers selected from the aforementioned procedure.

[0212] A multi-class prediction system based on (Friedman et al., "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, **33**(1): 1, 2010) was constructed to predict the group membership of samples in the test data using the panel of markers selected. A confusion matrix and ROC curves were also provided to evaluate sensitivity and specificity, in addition to prediction accuracy.

[0213] The hypothesis testing was two-sided with p-value ≤ 0.05 and was considered to be statistically significant. The analysis was conducted in R version 3.3.2 with the following packages used: 'glmnet', 'lpc', 'CoxBoost', 'limma', and 'ROCR'.

[0214] Table 1 shows a list of markers presented in at least 7 out of 10 random split. Based on 10 random split, the median prediction accuracy on test data was 98.0% with minimum 47.7% and maximum 99.1%. In some instances, this list of markers is used for diagnostic purposes.

Table 1.

Colon-cancer	"cg08088171" "cg13420112" "cg14642259" "cg20973720" "cg24583770" "cg27536151"
Colon-normal	"cg24741563" "cg22979615"

[0215] Table 2 shows the confusion matrix (or error matrix) on TCGA test dataset. The prediction is based on a list of markers presented in at least 7 out of 10 random split.

Table 2.

	Colon-cancer	**Colon-normal**
Colon-cancer	129	7
Colon-normal	0	11

Example 3. Prognosis Analysis

[0216] DNA methylation data were obtained from the TCGA analysis of about 27,000 sites generated using the Infinium 450K Methylation Array. Methylation profile for colon cancer tissue was analyzed. Four clinical covariates were used as described in Example 1.

[0217] The data were further modified by removal of 1) negative and 0 survival time; or 2) missing value in any of the four clinical and demographic covariates; to generate 365 colon cancer samples with 53 events (the event of interest was defined as death) for subsequent diagnostic analysis.

[0218] Additional datasets were also obtained from TCGA which includes:

[0219] Breast cancer: 967 observations with 103 events;

[0220] Liver cancer: 355 observations with 98 events;

[0221] Lung adenocarcinoma (LUAD): 516 observations with 141 events;

[0222] Lung squamous cell carcinoma (LUSC): 418 observations with 150 events;

[0223] and were used during the subsequent diagnostic analysis.

[0224] For each type of cancer: breast, colon, liver, and lung (LUAD and LUSC combined), the full dataset was split randomly into training and test sets with 2:1 ratio. A randomized lasso was used (Meinshausen and Bühlmann, "Stability selection," Journal of the Royal Statistical Society: Series B Statistical Methodology, 72(4): 417-473, 2010), the random split and variable selection scheme were repeated for 10 times, which provides finite sample error control and improves stability of variable selection. For each random split, the univariate pre-screening procedure was first performed on the training data to remove excessive noise and accelerate the computational procedure (Wasserman and Roeder, "High dimensional variable selection. Annals of statistics, 37(5A): 2178, 2009). For each methylation marker, a univariate Cox proportional hazards model was fitted by using each marker as the covariate. A marker with p-value ≤ 0.05 from the Wald statistic was retained in the dataset.

[0225] Four variable selection methods suitable for high-dimensionality on the prescreened training dataset were applied: Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), p267-288, 1996), Elastic Net (Zou and Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2): 301-320, 2005), Lassoed Principal Components (LPC) (Witten and Tibshirani, "Testing significance of features by lassoed principal components," The annals of applied statistics, 2(3): 986, 2008) and Boosting [Binder et al., "Boosting for high-dimensional time-to event data with competing risks," Bioinformatics, 25(7): 890-896, 2009]. For LASSO and elastic net, the tuning parameters (λ for LASSO and λ, α for elastic net) were determined according to the expected generalization error estimated from 10-fold cross-validation and information-based criteria AIC/BIC. For LPC, the markers with p-value ≤ 0.05 after False Discovery Rate (FDR) correction were considered to be statistically significant. For boosting, the optimal step was determined by expected generalization error estimated from 10-fold cross-validation. For all methods, the number of markers was also governed by the effective sample size in the training dataset, which approximately equaled to the number of events (death). A Cox proportional hazards model was fitted on the training data using markers selected at the optimal step as the covariates. The predictability of the model was evaluated by ρ^2 on the training data and concordance probability (also known as c-index) on the test data. ρ^2 - the proportion of explained randomness (O'Quigley et al., "Explained randomness in proportional hazards models," *Statistics in medicine*, 24(3): 479-489, 2005), is a function of Kullback- Leibler information gain bounded between 0 and 1, with a larger value indicating larger proportion of randomness explained. C-index (Harrell et al., "Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statistics in medicine*, 15: 361-387, 1996) calculates the proportions of concordant pairs among all pairs of observations with 1 indicating perfect prediction accuracy. The selection of the optimal tuning parameter was difficult given the variability in samples, as such, a composite panel was constructed by keeping markers with a high selection probability and disregard those with low selection probability from the 10 sets of markers selected from the aforementioned procedure.

[0226] To validate, a Cox proportional hazards model was fitted using the panel of markers with high selection probability on the training data. By multiplying the coefficient estimates and the design matrix in the test data, a risk score was obtained for each observation in the test data. By dividing the risk score according to its median, a high risk group and a low risk group were formed with roughly equal number of observations. A plot of Kaplan-Meier estimator and log-rank test were included to determine if the median survival time was significantly different. A model-based prediction was also provided. The high concordance between the non-parametric and semi-parametric prediction curves indicated the possibility of accurately predicting a new patient's survival status for any future time point using the panel of markers selected.

[0227] Table 3 illustrates a list of markers selected in 3 out of 10 training/test split. In some instances, the list of markers is used for prognosis of a subject having colon cancer. Table 4 illustrates validation and prediction performance. Table 5 shows a log-rank test on the test data (Chi-squared test = 0.7, df=1, p=0.414). High and low risk group classification is based on risk score using markers selected by the method LASSO. Table 6 illustrates a log-rank test on the test data (Chi-squared test = 0, df=1, p=0.954). High and low risk group classification is based on risk score using markers selected by the method Boosting.

Table 3.

LASSO	"cg13510262" "cg26952662"	"cg13620770"	"cg13859324"	"cg21685565"	"cg21922841"
Boosting	"cg13510262" "cg26952662"	"cg13859324"	"cg14159672"	"cg17675150"	"cg21685565"
Overlapping	"cg13510262"	"cg13859324"	"cg21685565"	"cg26952662"	

Table 4.

LASSO	ρ^2 on training c-index on test	0.87 (min: 0.43 - max: 0.99) 0.65 (min: 0.35 - max: 0.75)
Boosting	ρ^2 on training c-index on test	0.90 (min: 0.70 - max: 0.99) 0.64 (min: 0.43 - max: 0.83)
log-rank LASSO	1/10 p-value < 0.05	
log-rank Boosting	0/10 p-value < 0.05	

Table 5.

	N	Observed	Expected
High risk	48	6	4.78
Low risk	48	3	4.22

Table 6.

	N	Observed	Expected
High risk	48	4	3.92
Low risk	48	5	5.08

[0228] Fig. 1 illustrates Kaplan-Meier curves from LASSO (left panel) and Bosting (right panel).

[0229] Fig. 2 shows Cox proportional hazards regression prediction curves from LASSO (left panel) and Boosting (right panel).

[0230] Fig. 3 shows a multiclass ROC curve from LASSO.

Example 4. Diagnosis of colon cancer utilizing a cell-free DNA sample

[0231] Cell-free DNA sample was obtained from a QIAamp Circulating Nucleic Acid Kit. Methylation profile of biomarker cp10673833 (Cob-2) was used for the analysis.

[0232] The methylation data from a total of 115 patients pre and post treatment (chemotherapy or surgery) were used, with an average of 2 month follow-up time. Among them, 39 patients have progressive disease (PD), 34 patients have partial response (PR), 27 patients have stable disease (SD) and 15 patients have tumor removed by surgery, which were assume to have complete response (CR). Boxplot (Fig. 4) shows methylation change pre- and post-treatment in four groups with different response. ANOVA (F-test = 187.44, p-value < 0.001) (Table 7) indicates the mean methylation change is different across four groups. By using Tukey's post-hoc analysis (Table 8), a pairwise comparison was conducted between all possible combinations. The methylation change is different between any two response groups, except between patients who have treated by surgery and those have partial response to chemotherapy. In addition, the CR and PR groups were combined as the response group, and repeated the aforementioned analysis (Fig. 5). ANOVA (F-test = 283.4, p-value < 0.001)

(Table 9) indicates the mean methylation change is significantly different across three groups. By using Tukey's post-hoc analysis (Table 10), the methylation changes between all pairwise comparisons are different.

Table 7. Summary statistics mean (standard deviation) and ANOVA results for colon cancer

	Surgery (n = 15)	PR (n = 34)	SD (n = 27)	PD (n = 39)	F-statistics	p-value
Δ cfDNA	-0.29 (0.18)	-0.28 (0.14)	-0.00 (0.06)	0.36 (0.13)	187.442	<.001***

Table 8. Tukey's post-hoc analysis on pairwise comparison for colon cancer

	mean.diff	lower 95 % CI	upper 95 % CI	adjust p-value
PR-surgery	0.01	-0.09	0.12	0.99
SD-surgery	0.28	0.18	0.39	0.00
PD-surgery	0.65	0.55	0.75	0.00
SD-PR	0.27	0.19	0.36	0.00
PD-PR	0.64	0.56	0.72	0.00
PD-SD	0.37	0.29	0.45	0.00

Table 9. Summary statistics mean (standard deviation) and ANOVA results for colon cancer

	Response (n = 49)	SD (n = 27)	PD (n = 39)	F-statistics	p-value
Δ cfDNA	-0.28 (0.15)	-0.00 (0.06)	0.36 (0.13)	283.382	<.001***

Table 10. Tukey's post-hoc analysis on pairwise comparison for colon cancer

	mean.diff	lower 95 % CI	upper 95 % CI	adjust p-value
SD-Response	0.28	0.20	0.35	0.00
PD-Response	0.65	0.58	0.71	0.00
PD-SD	0.37	0.29	0.44	0.00

Example 5. Identification of Methylation Correlated Block (MCB)

[0233] In some instances, closely positioned CpG have similar methylation levels, due to a processivity and lack of sequence-specificity of DNA methyltransferases and demethylases, as well as the concept of haplotype blocks in genetic linkage analysis. Pearson correlation coefficients r2 between β values of any two CpGs positioned within one kilobase of one another were calculated. A cutoff of r2>0.5 was used to identify Methylation Correlated Block (MCB) (also refers to herein as BCM) within regions interrogated by the padlock probes. A value of Pearson's r<0.5 was used to identify transition spots (boundaries) between any two adjacent markers indicating uncorrelated methylation. Markers not separated by a boundary were combined into Methylation Correlated Block (MCB). This procedure combined between 2 and 22 CpG positions in each block to identify a total number of BCMs in each diagnostic category within the padlock data. Methylation frequencies for entire MCBs were calculated by summing up the numbers of Cs at all interrogated CpG positions within a BCM and dividing by the total number of C+Ts at those positions Pearson correlation coefficients between methylation frequencies of each pair of CpG markers separated by no more than 200bp were calculated separately from 30 cancer and 30 corresponding normal tissue samples from each of the two diagnostic categories. A value of Pearson's r<0.5 was used to identify transition spots (boundaries) between any two adjacent markers indicating uncorrelated methylation. Markers not separated by a boundary were combined into Methylation Correlated Block (MCB). Methylation frequencies for entire BCMs were calculated by summing up the numbers of Cs at all interrogated CpG positions within a BCM and dividing by the total number of C+Ts at those positions.

[0234] Fig. 6 illustrates an exemplary BCM for colon cancer. Data was compiled from both healthy tissue samples and corresponding cancers. Row I represents a CG dinucleodide analyzed in this study. Subset of those correspond to CG markers included on Illumina's Infinium HumanMethylation450 BeadChip (leftmost column), whereas the majority represent previously unknown potential novel markers. Row II lists genomic distances between markers and Row III shows genomic location of each analyzed CG. Each row within Row I illustrates Pearson correlation coefficients r2 between β values of two closely positioned CGs calculated for samples from each tissue type separately. Correlation between any two markers is represented by a gray square at the intersection of (virtual) perpendicular lines originating from these two markers. White color indicates no significant correlation, gray color intensity marks r2 values between 0.5 and 1. Black boxes indicate the ends of analyzed regions.

Example 6. Linking differentially methylated markers to gene expression

[0235] TCGA DNA methylation and RNAseq expression data for colon cancer samples are obtained from the TCGA website. The degree of DNA methylation at each CpG is denoted as a beta value and is calculated as (M/(M+U)), where M and U are normalized values representing the methylated and unmethylated allele intensities respectively. Beta values range from 0 to 1 and reflect the fraction of methylated alleles at each CpG in each sample. The methylation beta value is calculated for all 485,000 markers for each of the breast tumor and matched normal breast tissues in the TCGA data. CpG markers with a mean value less than .05 or greater than .95 are selected for further evaluation. Markers with a difference between the mean methylation value for the tumor tissue and the mean methylation value of the corresponding normal tissue of greater than 0.5 are also selected. At the intersection of these two groups, markers for which the mean methylation is <.05 for normal colon tissue samples and the difference between normal and tumor is greater than 0.5 are further selected and the genes associated with these markers are identified. For each marker, the tumor samples are then separated into those with methylation values greater than the mean value of the tumor samples and those with methylation values less than the mean value of the tumor samples. Next, the RNAseq data in the TCGA data is examined and the relative expression of each gene is calculated. Because of the wide variation of the expression values, the values are adjusted as follows: log2 (expressionValue + 1). Genes are identified in which the difference in the methylation values correlated with variation in the associated gene expression levels. Genes for which there is a correlation were selected for further functional evaluation and validation.

[0236] DNA/RNA isolation and Quantitative PCR

[0237] Tumor and corresponding far site samples of the same tissue are obtained from patients who underwent surgical tumor resection; samples are frozen and preserved at -80^{O}C until use. Isolation of DNA and RNA from samples is performed using AllPrep DNA/RNA Mini kit (Qiagen, Valencia, CA) according to the manufacturer's recommendations. During RNA isolation, the sample is subjected to on-column DNase digestion. RNA is quantified using a Nanodrop 2000 (Thermo Scientific). 200ng RNA of each sample is used for cDNA synthesis using iScript cDNA synthesis kit (Bio-rad, Inc) according to the manufacturer's instructions. qPCR is performed by a standard 40-cycle amplification protocol using gene-specific primers and a Power SYBR Green PCR Master Mix on a 7500 Real Time PCR system (Applied Biosystems). Experiments are carried out in triplicate and normalized to endogenous *ACTB* levels. Relative fold change in expression is calculated using the ΔΔCT method (cycle threshold values <30).

[0238] Cell culture and gene transfections

[0239] Human colon cancer cell line HCT-116 and the human embryonic kidney cell line HEK293a are obtained from American type culture collection (Manassas, VA, USA) and cultured according to their instructions. The expression construct for a gene disclosed herein is purchased from Origene in a form of TrueORF ® cDNA clones in pCMV6-Entry vector. cDNAs are shuttled into pLenti-C-mGFP (Origene) to create a lentivector encoding a fusion protein between the desired gene and mGFP.

[0240] Lentiviral particles are made by co-transfection of HEK-293T cells with a colon cancer gene mGFP lentivector together with a third-generation packaging vector using calcium phosphate precipitation. Viral supernatants are collected 36 hours post-transfection. Human colon cancer cell line HCT-116 are plated in a 6-well plate the day before transfection of the breast cancer gene mGFP lentivector. Stable cell lines are generated by infecting cells with viral particles at the MOI of ~5 for 24hrs and are collected and sorted to 100% GFP – positivity using FACS. The GFP positive cells are then used for a colony formation assay in cell culture and tumor xenograft in nude mice.

[0241] Colony formation assays

[0242] Cells are plated in 6-well plates at a density of 500 cells per well and are cultured at 37°C with 5% CO2 humidified air for 14 days. The colonies are fixed with 10% formaldehyde for 5 min and then stained with 0.1% crystal violet for 30 seconds. Colony consisting of 50 or more cells are counted. The experiment is performed in triplicate and repeated 3 times. Plate efficiency = (colony numbers /inoculated cell numbers) × 100%.

[0243] Tumor xenograft

[0244] All animal studies are performed in accordance with institutional and international animal regulations. Animal protocols are approved by the Institutional Animal Care and Use Committee of Sun Yat- Sen University. Female athymic BALB/c nude mice (4–5 weeks of age, 18–20 g) are purchased from a vendor (Guangdong Province Laboratory Animal Center, Guangzhou, China).

[0245] Mice are injected subcutaneously with 100 µl of tumor cells suspended in serum free medium. Tumor growth is monitored every 3 days by visual examination. Tumor sizes are measured using a caliper, and tumor volume is calculated according to the following equation: tumor volume (mm3) = (length (mm) × width (mm)2) × 0.5. All animals are sacrificed 3-4 weeks postinjection and the xenografts were harvested. Representative data are obtained from five mice per experimental group. Statistical analyses are performed with one-way repeated- measures ANOVA.

Example 7.

[0246] Table 11A and Table 11B illustrate the gene names referenced by the CpG sites described herein. Table 11A illustrates the CpG site and the respective gene name.

CpG Site	Gene Name	Description	Alternatives
cg08088171	CANT1	calcium activated nucleotidase 1	soluble Ca-activated nucleotidase, isozyme 1; SCAN-1; SHAPY; and DBQD

CpG Site	Gene Name	Description	Alternatives
cg27536151	MAP2K5	mitogen-activated protein kinase kinase 5	MAP kinase kinase 5; MAPK/ERK kinase 5; MEK5; and MKK5
cg13420112	ZNF512B	Zinc Finger Protein 512B	KIAA1196; GM632
cg14642259	MYBPC3	Myosin Binding Protein C, Cardiac	CMD1MM; LVNC10; CMH4
cg20973720	RASGRF1	Ras protein specific guanine nucleotide releasing factor 1	Ras-specific nucleotide exchange factor CDC25; GNRP; CDC25L; and GRF55
cg24741563	APPBP2	Amyloid Beta Precursor Protein Binding Protein 2	Protein Interacting With APP Tail 1; PAT1; and KIAA0228
cg22979615	HCCA2	MOB kinase activator 2	Mob2 homolog
cg10673833	MYO1G	myosin IG	minor histocompatibility antigen HA-2; unconventional myosin-Ig; HLA-HA2; MHAG
cg13510262	SP1	Sp1 transcription factor	Specific protein 1 and TSFP1
cg13859324	UNC45B	Unc-45 Myosin Chaperone B	Cardiomyopathy Associated 4; CTRCT43
cg26952662	CTHRC1	Collagen Triple Helix Repeat Containing 1	Protein NMTC1
cg14159672	--	--	--
cg17675150	ZNF532	Zinc Finger Protein 532	KIAA1629
cg13620770	GPR137	G Protein-Coupled Receptor 137	Transmembrane 7 Superfamily Member 1-Like 1 Protein; C11orf4
cg21922841	SLC9A3R1	SLC9A3 Regulator 1	NHERF-1; EBP50; NPHLOP2

[0247] Table 11B illustrates the associated chromosome location with the CpG sites described herein.

CpG Site	Genome _Build	CHR	MAPINFO	Chromosome_36	Coordinate_36	Gene Name
cg08088171	37	17	7.7E+07	17	74514708	CANT1
cg27536151	37	15	6.8E+07	15	65651407	MAP2K5
cg13420112	37	20	6.3E+07	20	62071098	ZNF512B
cg14642259	37	11	4.7E+07	11	47315593	MYBPC3
cg20973720	37	15	7.9E+07	15	77170222	RASGRF1
cg24741563	37	17	5.9E+07	17	55900117	APPBP2
cg22979615	37	11	1510644	11	1467220	HCCA2
cg10673833	37	7	4.5E+07	7	44985374	MYO1G
cg13510262	37	12	5.4E+07	12	52060304	SP1
cg13859324	37	17	3.3E+07	17	30498805	UNC45B
cg26952662	37	8	1E+08	8	104452675	CTHRC1
cg14159672	37	1	2.1E+08	1	204085802	--
cg17675150	37	18	5.7E+07	18	54680764	ZNF532
cg13620770	37	11	6.4E+07	11	63808314	GPR137
cg21922841	37	17	7.3E+07	17	70255726	SLC9A3R1

Example 8.

[0248] Table 12 illustrates padlock probes described herein.

CpG Site	Gene	Probe Sequence	SEQ ID NO:
cg08088171	CANT1	CCTTCTTCACTTCCCATAAAACCATCCCCTGTCTCT TATACACATCTCCGAGCCCACGAGACCCCCCCCCC CCCCCTCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNNNNAAAAATCCCAATCTCACCTCTA	1
cg27536151	MAP2K5	ATTTTTCTTATCTTCCCACACTAAACTCTCTGTCTCT TATACACATCTCCGAGCCCACGAGACCCCCCCCTC GTCGGCAGCGTCAGATGTGTATAAGAGACAGNNN NNNACCTTTTTAATCCCCAAAACTAAATTAA	2
cg13420112	ZNF512B	CTCTACCTCCCAAACTAACCCTAACCTCAAACCTGT CTCTTATACACATCTCCGAGCCCACGAGACCCCCC CCCCCCCCCCCCTCGTCGGCAGCGTCAGATGTGTAT AAGAGACAGNNNNNNAAACCCCCTCTCCCC	3
cg14642259	MYBPC3	TTCATAACTCAACTCCTAAATCAAATCCTGTCTCTT ATACACATCTCCGAGCCCACGAGACCCCCCCCCCC CCCCCCCCTCGTCGGCAGCGTCAGATGTGTATAAG AGACAGNNNNNNAACTAAACCTAAACATACC	4
cg20973720	RASGRF1	CAATCCCTCAATTTACCCCTGTCTCTTATACACATC TCCGAGCCCACGAGACCCCCCCCCCCCCCCCCCCC CCCCCCCCTCGTCGGCAGCGTCAGATGTGTATAAG AGACAGNNNNNNAAAAACACCAAATTACAAC	5
cg24741563	APPBP2	ACTTATCTCTCCCTTCTCATTTCTCTTTACACTGTCT CTTATACACATCTCCGAGCCCACGAGACCCCCCCC CTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGN NNNNNATTAAATAATACATAAATCAAATCA	6
cg22979615	HCCA2	ATTTCCCTCTATACTCTTCCTACCCTGTCTCTTATAC ACATCTCCGAGCCCACGAGACCCCCCCCCCCCCCC CCCCCTCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNNNNAATCACAACCTCAACCAAAAA	7
cg10673833	MYO1G	AACACAACCTCCTTATAAAACCTGTCTCTTATACAC ATCTCCGAGCCCACGAGACTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGNNNNNNAACIAAAAACCC TCCAAA	8
cg13510262	SP1	ACAAAACCCCCTTAAACTTATCCCTCAACTCTGTCT CTTATACACATCTCCGAGCCCACGAGACCCCCCCC CCCCCCCCCTCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNTCCCTCCTCCTTACCCCC	9
cg13859324	UNC45B	TCAATAACAAACTCCAAAACTGTCTCTTATACACA TCTCCGAGCCCACGAGACCCCCCCCCCCCCCCCCC CCCCCCCCCCCCTCGTCGGCAGCGTCAGATGTGTAT AAGAGACAGNNNNNNCCCCACCAAAATAACCC	10
cg26952662	CTHRC1	CCCAAACCCTACAAACCTGTCTCTTATACACATCTC CGAGCCCACGAGACCCCCCCCCCCCCCCCCCCCCC CCCCCCCCTCGTCGGCAGCGTCAGATGTGTATAAG AGACAGNNNNNNTATACCCAAATTAAACCAC	11
cg14159672	--	ACCCTAATAACTATACTACTCCTAATTTCTGTCTCT TATACACATCTCCGAGCCCACGAGACCCCCCCCCC CCCCCCCCTCGTCGGCAGCGTCAGATGTGTATAAG AGACAGNNNNNNTTTAAATCAAAACTACCCC	12
cg17675150	ZNF532	AACCAAACAAAAACCCCCAACTGTCTCTTATACAC	13

		ATCTCCGAGCCCACGAGACCCCCCCCCCCCCCCC CCCCCTCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNNNNAAATAAAATATACTATTTTCCCA	
cg13620770	GPR137	ACTAAAACCCAAACCCCTCTCTCTGTCTCTTATACA CATCTCCGAGCCCACGAGACCCCCCCCCCCCCCC CCCCCCTCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGNNNNNNTACTAATTAACTACTCCTACT	14
cg21922841	SLC9A3R1	ACAACCTTTAACCAAAACCTAAATACAACTCCCTG TCTCTTATACACATCTCCGAGCCCACGAGACCCCCC CCCCCCCCCTCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNTTTCAAACTCCTAACCTC	15

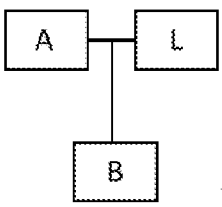
[0249] Embodiment 1 refers to a method of generating a methylation profile of a biomarker in a subject in need thereof, comprising: (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject; (b) detecting a hybridization between the extracted genomic DNA and a probe, wherein the probe hybridizes to a biomarker selected from cg08088171 (calcium activated nucleotidase 1), cg27536151 (mitogen-activated protein kinase kinase 5), cg13420112 (zinc finger protein 512B), cg14642259 (myosin binding protein C, cardiac), cg20973720 (Ras protein specific guanine nucleotide releasing factor 1), cg24741563 (amyloid beta precursor protein binding protein 2), cg22979615 and cg10673833 (myosin IG); and (c) generating a methylation profile based on the detected hybridization between the extracted genomic DNA and the probe.

[0250] Embodiment 2: the method of embodiment 1, wherein the methylation profile comprises cg10673833.

[0251] Embodiment 3: the method of embodiment 1, wherein the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563 and cg22979615.

[0252] Embodiment 4: the method of embodiment 1, wherein the methylation profile comprises cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563 and cg22979615.

[0253] Embodiment 5: the method of embodiment 1, wherein the probe comprises a structure of Formula I:



Formula I

wherein:

A is a first target-binding region;

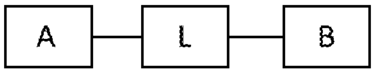
B is a second target-binding region; and

L is a linker region;

wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 30 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15; B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 12 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15; and

wherein L is attached to A; and B is attached to either A or L.

[0254] Embodiment 6: the method of embodiment 5, wherein the probe comprises a structure of Formula Ia:



Formula Ia.

[0255] Embodiment 7: the method of embodiment 5, wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 40 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15.

[0256] Embodiment 8: the method of embodiment 5, wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 50 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15.

[0257] Embodiment 9: the method of embodiment 5, wherein B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 15 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15.

[0258] Embodiment 10: the method of embodiment 5, wherein B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 20 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15.

[0259] Embodiment 11: the method of embodiment 5, wherein L is about 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 nucleotides in length.

[0260] Embodiment 12: the method of embodiment 1, wherein the probe comprises at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a sequence selected from SEQ ID NOs: 1-15.

[0261] Embodiment 13: the method of embodiment 1, wherein the generating further comprises generating a pair-wise methylation difference dataset comprising: (i) a first difference between the methylation profile of the treated genomic DNA with a methylation profile of a first normal sample; (ii) a second difference between a methylation profile of a second normal sample and a methylation profile of a third normal sample; and (iii) a third difference between a methylation profile of a first primary cancer sample and a methylation profile of a second primary cancer sample.

[0262] Embodiment 14: the method of embodiment 13, wherein the generating further comprises analyzing the pair-wise methylation difference dataset with a control by a machine learning method to generate the methylation profile.

[0263] Embodiment 15: the method of embodiment 13, wherein the first primary cancer sample is a colon cancer sample.

[0264] Embodiment 16: the method of embodiment 13, wherein the second primary cancer sample is a non-colon cancer sample.

[0265] Embodiment 17: the method of embodiment 14, wherein the control comprises a set of methylation profiles, wherein each said methylation profile is generated from a biological sample obtained from a known cancer type.

[0266] Embodiment 18: the method of embodiment 17, wherein the known cancer type is colon cancer.

[0267] Embodiment 19: the method of embodiment 17, wherein the known cancer type is a relapsed or refractory colon cancer.

[0268] Embodiment 20: the method of embodiment 17, wherein the known cancer type is a metastatic colon cancer.

[0269] Embodiment 21: the method of embodiment 17, where the known cancer type is adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma.

[0270] Embodiment 22: the method of embodiment 14, wherein the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.

[0271] Embodiment 23: the method of embodiment 1, wherein the method further comprises performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers prior to generating the methylation profile.

[0272] Embodiment 24 refers to a method of selecting a subject suspected of having colon cancer for treatment, the method comprising: (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having colon cancer; (b) generating a methylation profile comprising one or more biomarkers selected from: cg08088171 (calcium activated nucleotidase 1), cg27536151 (mitogen-activated protein kinase kinase 5), cg13420112 (zinc finger protein 512B), cg14642259 (myosin binding protein C, cardiac), cg20973720 (Ras protein specific guanine nucleotide releasing factor 1), cg24741563 (amyloid beta precursor protein binding protein 2), cg22979615, and cg10673833 (myosin IG) from the extracted genomic DNA; (c) comparing the methylation profile of the one or more biomarkers with a control; (d) identifying the subject as having colon cancer if the methylation profile correlates to the control; and (e) administering an effective amount of a therapeutic agent to the subject if the subject is identified as having colon cancer.

[0273] Embodiment 25: the method of embodiment 24, wherein the methylation profile comprises cg10673833.

[0274] Embodiment 26: the method of embodiment 24, wherein the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563 and cg22979615.

[0275] Embodiment 27: the method of embodiment 24, wherein the methylation profile comprises cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563 and cg22979615.

[0276] Embodiment 28: the method of embodiment 24, wherein the comparing further comprises generating a pair-wise methylation difference dataset comprising: (i) a first difference between the methylation profile of the treated genomic DNA with a methylation profile of a first normal sample; (ii) a second difference between a methylation profile of a second normal sample and a methylation profile of a third normal sample; and (iii) a third difference between a methylation profile of a first primary cancer sample and a methylation profile of a second primary cancer sample.

[0277] Embodiment 29: the method of embodiment 28, wherein the comparing further comprises analyzing the pair-wise methylation difference dataset with a control by a machine learning method to generate the methylation profile.

[0278] Embodiment 30: the method of embodiment 28, wherein the first primary cancer sample is a colon cancer sample.

[0279] Embodiment 31: the method of embodiment 28, wherein the second primary cancer sample is a non-colon cancer sample.

[0280] Embodiment 32: the method of embodiment 29, wherein the control comprises a set of methylation profiles, wherein each said methylation profile is generated from a biological sample obtained from a known cancer type.

[0281] Embodiment 33: the method of embodiment 32, wherein the known cancer type is colon cancer.

[0282] Embodiment 34: the method of embodiment 32, wherein the known cancer type is a relapsed or refractory colon cancer.

[0283] Embodiment 35: the method of embodiment 32, wherein the known cancer type is a metastatic colon cancer.

[0284] Embodiment 36: the method of embodiment 29, wherein the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.

[0285] Embodiment 37: the method of embodiment 24, wherein the generating further comprises hybridizing each of the one or more biomarkers with a probe, and performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers.

[0286] Embodiment 38 refers to a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in the subject, comprising: (a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject having colon cancer; (b) generating a methylation profile comprising one or more biomarkers selected from: cg13510262 (Sp1 transcription factor), cg13859324 (Unc-45 myosin chaperone B), and cg26952662 (collagen triple helix repeat containing 1) from the extracted genomic DNA; (c) obtaining a methylation score based on the methylation profile of the one or more biomarkers; and (d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission, initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent.

[0287] Embodiment 39: the method of embodiment 38, wherein the methylation profile comprises cg13510262, cg13859324 and cg26952662.

[0288] Embodiment 40: the method of embodiment 38, wherein the methylation profile further comprises one or more biomarkers selected from: cg14159672, cg17675150 (zinc finger protein 532), cg13620770 (G protein-coupled receptor 137), and cg21922841 (SLC9A3 regulator 1).

[0289] Embodiment 41: the method of embodiment 38, wherein the methylation profile further comprises one or more biomarkers selected from: cg13620770 and cg21922841.

[0290] Embodiment 42: the method of embodiment 38, wherein the methylation profile further comprises one or more biomarkers selected from: cg14159672 and cg17675150.

[0291] Embodiment 43: the method of embodiment 38, wherein the methylation profile further comprises cg14159672, cg17675150, cg13620770, and cg21922841.

[0292] Embodiment 44: the method of embodiment 38, wherein the methylation score of from about 1.5 to about 3 is indicative of a survival for at least 6 months, at least 1 year, at least 1.5 years, at least 2 years, at least 2.5 years, at least 3 years, at least 4 years, or at least 5 years.

[0293] Embodiment 45: the method of embodiment 38, wherein the methylation score of from about 1.5 to about 2.5 is indicative of a survival for at least 6 months, at least 1 year, at least 1.5 years, at least 2 years, at least 2.5 years, at least 3 years, at least 4 years, or at least 5 years.

[0294] Embodiment 46: the method of embodiment 38, wherein the methylation score of less than 1.5 is indicative of a survival of less than 5 years, less than 4 years, less than 3 years, less than 2.5 years, less than 2 years, less than 1.5 years, less than 1 year, or less than 6 months.

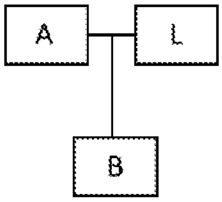
[0295] Embodiment 47: the method of embodiment 38, wherein the methylation score is calculated based on Cox proportional hazards (PH) regression analysis.

[0296] Embodiment 48: the method of any one of the embodiments 38-47, wherein colon cancer is metastatic colon cancer.

[0297] Embodiment 49: the method of any one of the embodiments 38-48, where colon cancer is adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma.

[0298] Embodiment 50: the method of embodiment 38, wherein the generating further comprises hybridizing each of the one or more biomarkers with a probe, and performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers.

[0299] Embodiment 51: the method of embodiment 37 or 50, wherein the probe comprises a structure of Formula I:



Formula I

wherein:

A is a first target-binding region;

B is a second target-binding region; and

L is a linker region;

wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 30 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15; B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 12 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15; and

wherein L is attached to A; and B is attached to either A or L.

[0300] Embodiment 52: the method of embodiment 51, wherein the probe comprises a structure of Formula Ia:

Formula Ia.

[0301] Embodiment 53: the method of embodiment 51, wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 40 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15.

[0302] Embodiment 54: the method of embodiment 51, wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 50 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15.

[0303] Embodiment 55: the method of embodiment 51, wherein B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 15 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15.

[0304] Embodiment 56: the method of embodiment 51, wherein B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 20 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15.

[0305] Embodiment 57: the method of embodiment 51, wherein L is about 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 nucleotides in length.

[0306] Embodiment 58: the method of embodiment 37 or 50, wherein the probe comprises at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a sequence selected from SEQ ID NOs: 1-15.

[0307] Embodiment 59: the method of any one of the embodiments 1-58, wherein the biological sample comprises a blood sample.

[0308] Embodiment 60: the method of any one of the embodiments 1-58, wherein the biological sample comprises a tissue biopsy sample.

[0309] Embodiment 61: the method of any one of the embodiments 1-58, wherein the biological sample comprises circulating tumor cells.

[0310] Embodiment 62: the method of any one of the embodiments 1-61, wherein the subject is a human.

[0311] Embodiment 63 refers to a kit comprising a set of nucleic acid probes that hybridizes to biomarkers: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563, cg22979615, and cg10673833.

[0312] Embodiment 64 refers to a kit comprising a set of nucleic acid probes that hybridizes to biomarkers: cg13510262, cg13859324 and cg26952662.

[0313] Embodiment 65: the kit of embodiment 64, wherein the kit further comprises a nucleic acid probe that hybridizes to a biomarker selected from cg14159672, cg17675150, cg13620770, and cg21922841.

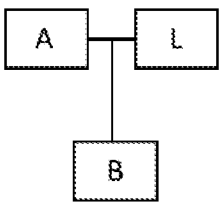
[0314] Embodiment 66: the kit of any one of the embodiments 63-65, wherein the set of nucleic acid probes comprises a set of probes selected from SEQ ID NOs: 1-15.

[0315] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein may be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A method of generating a methylation profile of a biomarker in a subject in need thereof, comprising:
   a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject;
   b) detecting a hybridization between the extracted genomic DNA and a probe, wherein the probe hybridizes to a biomarker selected from cg08088171 (calcium activated nucleotidase 1), cg27536151 (mitogen-activated protein kinase kinase 5), cg13420112 (zinc finger protein 512B), cg14642259 (myosin binding protein C, cardiac), cg20973720 (Ras protein specific guanine nucleotide releasing factor 1), cg24741563 (amyloid beta precursor protein binding protein 2), cg22979615 and cg10673833 (myosin IG); and
   c) generating a methylation profile based on the detected hybridization between the extracted genomic DNA and the probe.
2. The method of claim 1, wherein the methylation profile comprises cg10673833.
3. The method of claim 1, wherein the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563 and cg22979615.
4. The method of claim 1, wherein the methylation profile comprises cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563 and cg22979615.
5. The method of claim 1, wherein the probe comprises a structure of Formula I:



Formula I

wherein:

A is a first target-binding region;

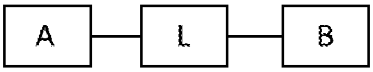
B is a second target-binding region; and

L is a linker region;

wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 30 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15; B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 12 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15; and

wherein L is attached to A; and B is attached to either A or L.

6. The method of claim 5, wherein the probe comprises a structure of Formula Ia:



Formula Ia.

7. The method of claim 5, wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 40 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15.

8. The method of claim 5, wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 50 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-15.

9. The method of claim 5, wherein B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 15 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15.

10. The method of claim 5, wherein B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 20 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-15.

11. The method of claim 5, wherein L is about 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 nucleotides in length.

12. The method of claim 1, wherein the probe comprises at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a sequence selected from SEQ ID NOs: 1-15.

13. The method of claim 1, wherein the generating further comprises generating a pair-wise methylation difference dataset comprising:

   (i) a first difference between the methylation profile of the treated genomic DNA with a methylation profile of a first normal sample;

   (ii) a second difference between a methylation profile of a second normal sample and a methylation profile of a third normal sample; and

(iii) a third difference between a methylation profile of a first primary cancer sample and a methylation profile of a second primary cancer sample.

14. The method of claim 13, wherein the generating further comprises analyzing the pair-wise methylation difference dataset with a control by a machine learning method to generate the methylation profile.
15. The method of claim 13, wherein the first primary cancer sample is a colon cancer sample.
16. The method of claim 13, wherein the second primary cancer sample is a non-colon cancer sample.
17. The method of claim 14, wherein the control comprises a set of methylation profiles, wherein each said methylation profile is generated from a biological sample obtained from a known cancer type.
18. The method of claim 17, wherein the known cancer type is colon cancer.
19. The method of claim 17, wherein the known cancer type is a relapsed or refractory colon cancer.
20. The method of claim 17, wherein the known cancer type is a metastatic colon cancer.
21. The method of claim 17, where the known cancer type is adenocarcinoma, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors, primary colorectal lymphoma, leiomyosarcoma, or melanoma and squamous cell carcinoma.
22. The method of claim 14, wherein the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.
23. The method of claim 1, wherein the method further comprises performing a DNA sequencing reaction to quantify the methylation of each of the one or more biomarkers prior to generating the methylation profile.
24. A method of selecting a subject suspected of having colon cancer for treatment, the method comprising:
    a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having colon cancer;
    b) generating a methylation profile comprising one or more biomarkers selected from: cg08088171 (calcium activated nucleotidase 1), cg27536151 (mitogen-

activated protein kinase kinase 5), cg13420112 (zinc finger protein 512B), cg14642259 (myosin binding protein C, cardiac), cg20973720 (Ras protein specific guanine nucleotide releasing factor 1), cg24741563 (amyloid beta precursor protein binding protein 2), cg22979615, and cg10673833 (myosin IG) from the extracted genomic DNA;

c) comparing the methylation profile of the one or more biomarkers with a control;
d) identifying the subject as having colon cancer if the methylation profile correlates to the control; and
e) administering an effective amount of a therapeutic agent to the subject if the subject is identified as having colon cancer.

25. A method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in the subject, comprising:

a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject having colon cancer;
b) generating a methylation profile comprising one or more biomarkers selected from: cg13510262 (Sp1 transcription factor), cg13859324 (Unc-45 myosin chaperone B), and cg26952662 (collagen triple helix repeat containing 1) from the extracted genomic DNA;
c) obtaining a methylation score based on the methylation profile of the one or more biomarkers; and
d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission, initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent.

26. The method of any one of the claims 1-25, wherein the biological sample comprises a blood sample.

27. The method of any one of the claims 1-25, wherein the biological sample comprises a tissue biopsy sample.

28. The method of any one of the claims 1-25, wherein the biological sample comprises circulating tumor cells.

Fig. 1

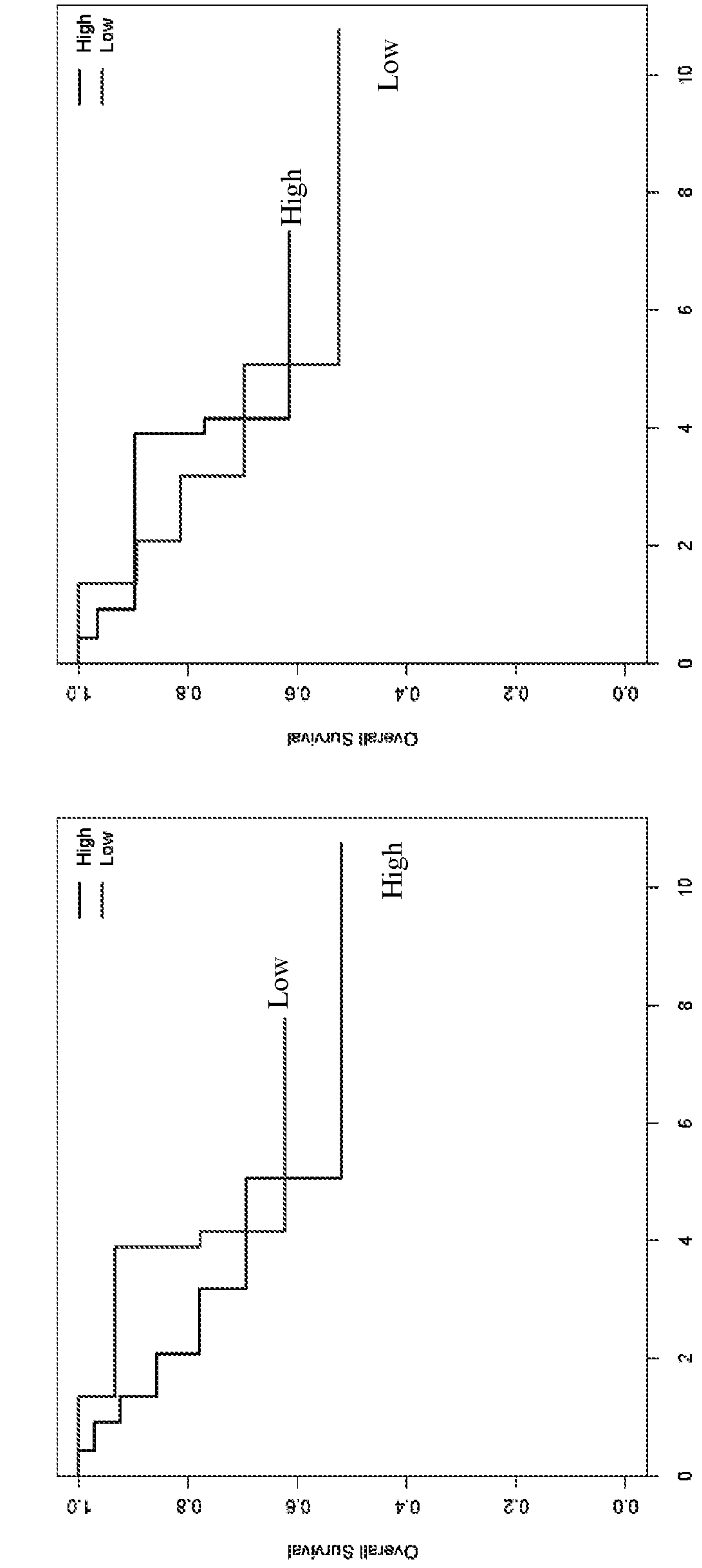
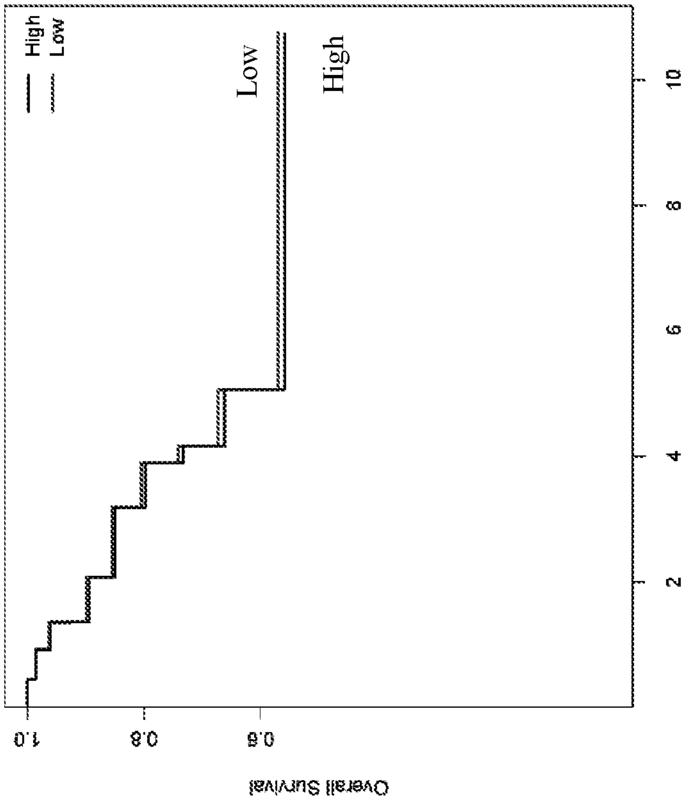

High
Low
Overall Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
High
Low
High
Low
Overall Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Low
High

Fig. 2


High
Low
Low
High
Overall Survival
1.0
0.8
0.6
2
4
6
8
10

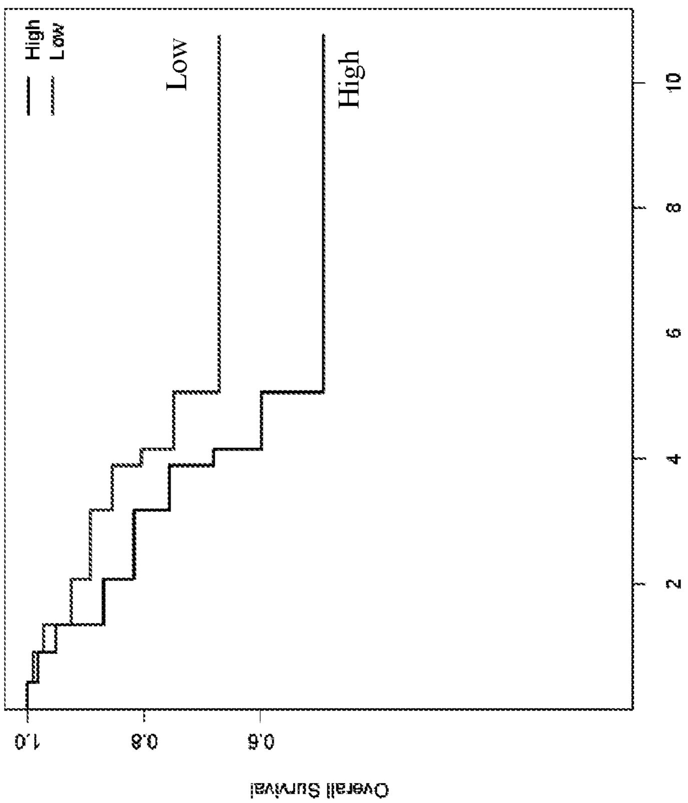

High
Low
Low
High
Overall Survival
1.0
0.8
0.6
2
4
6
8
10

Fig. 3

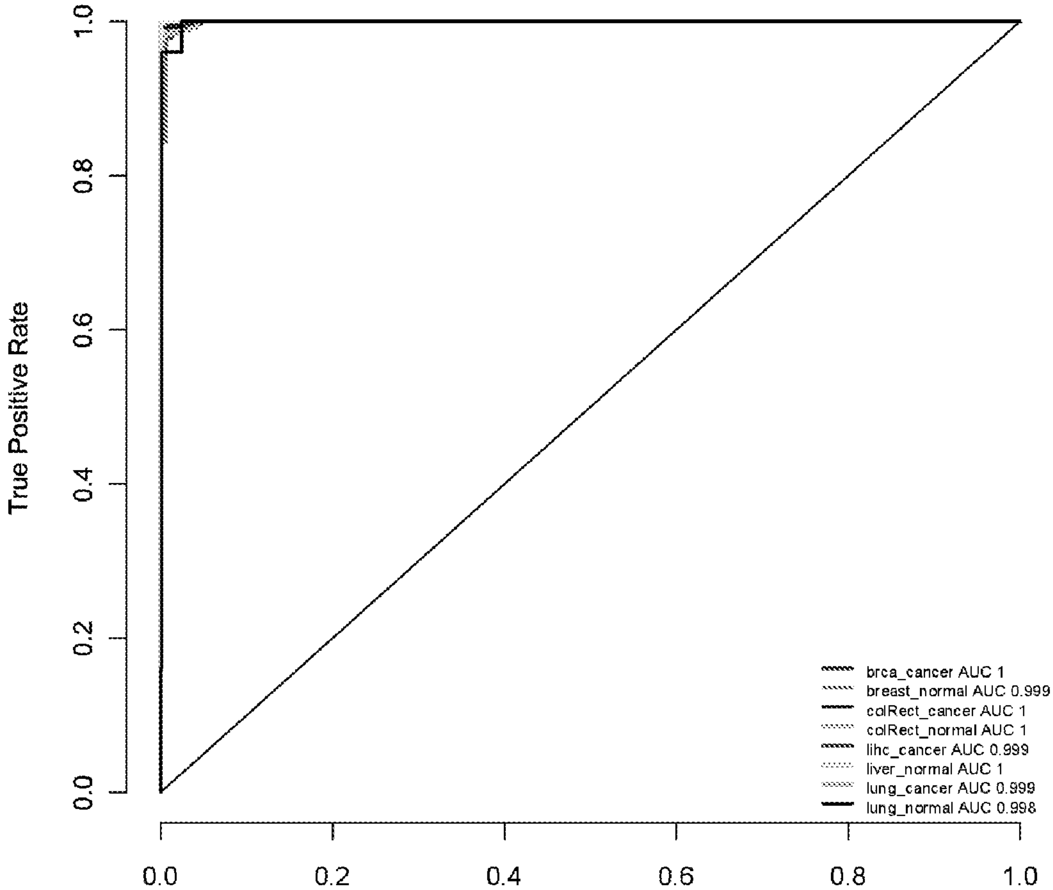

1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
brca_cancer AUC 1
breast_normal AUC 0.999
colRect_cancer AUC 1
colRect_normal AUC 1
lihc_cancer AUC 0.999
liver_normal AUC 1
lung_cancer AUC 0.999
lung_normal AUC 0.998
0.0
0.2
0.4
0.6
0.8
1.0

Fig. 4

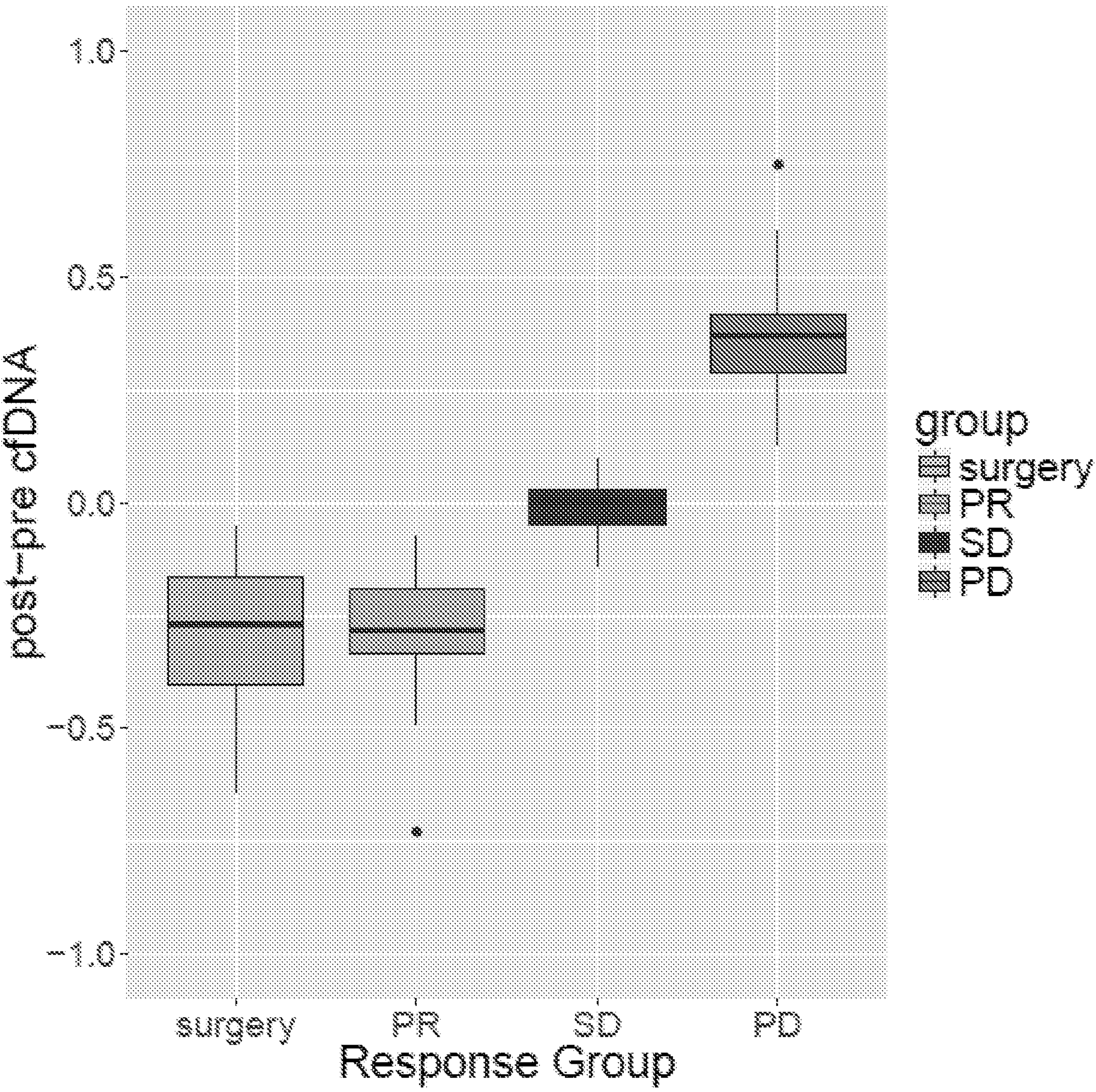

1.0
0.5
0.0
-0.5
-1.0
post-pre cfDNA
surgery
PR
SD
PD
Response Group
group
surgery
PR
SD
PD

Fig. 5

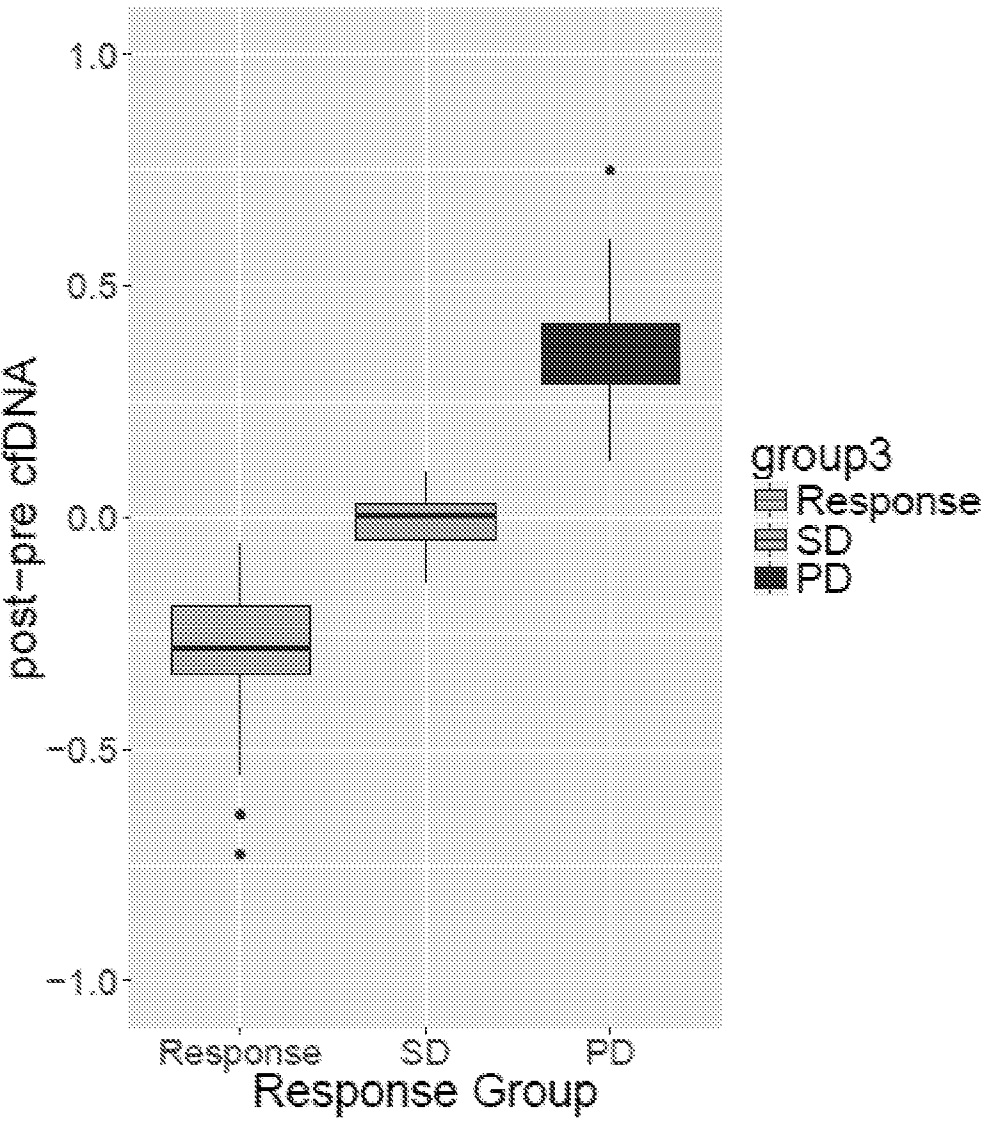
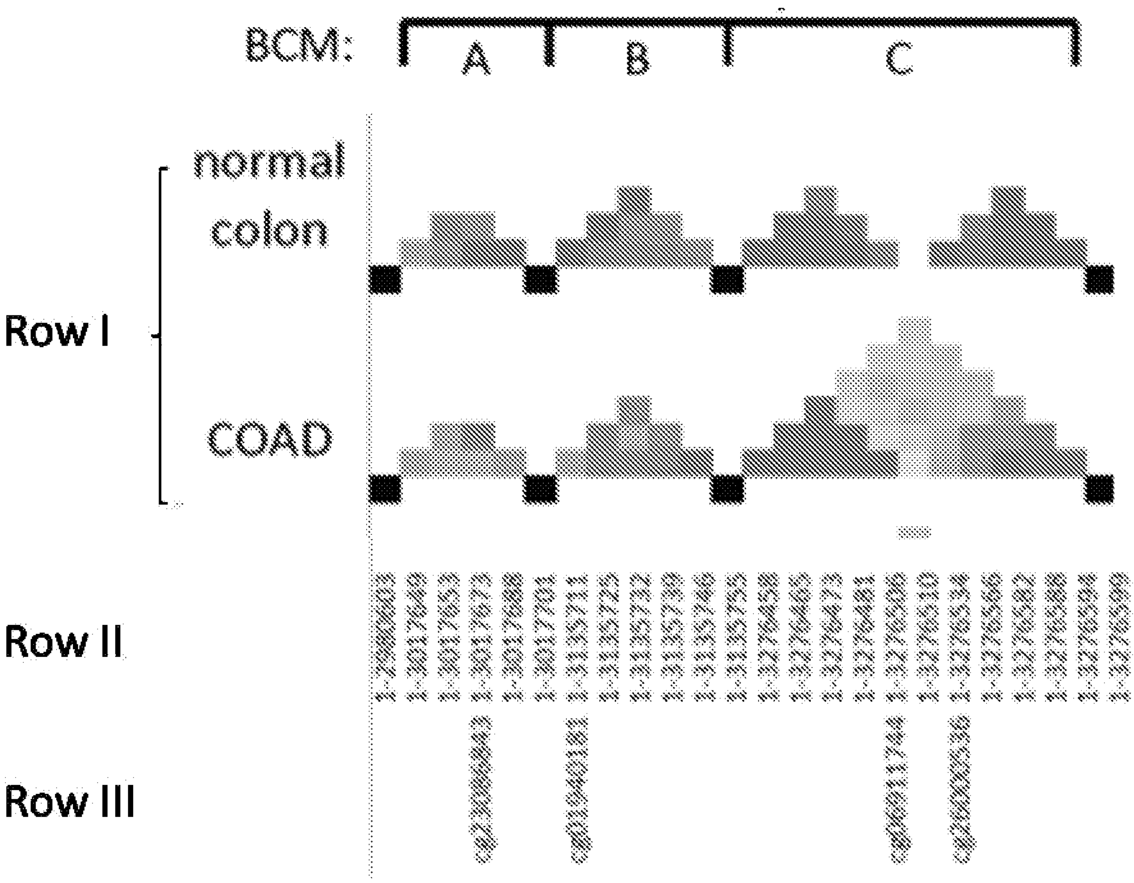

1.0
0.5
0.0
-0.5
-1.0
post-pre cfDNA
Response
SD
PD
Response Group
group3
Response
SD
PD

Fig. 6



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:

SEQ ID NOs:1-15 were searched.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/040945

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.:

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.:
1, 3, 5-24, and 26-28 to the extent that they read on a biomarker of cg08088171 and a probe of the structure of Formula Ia, where A is SEQ ID NO:1, B is SEQ ID NO:2, and L is a linker of 15 nucleotides in length.

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.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/040945

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 15/11; C12Q 1/68; G01N 33/574 (2017.01)
CPC - C12Q 1/6886; C12Q 2521/331; C12Q 2600/112; C12Q 2600/154; C12Q 2600/158 (2017.08)

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
USPC - 435/6.11; 506/9 (keyword delimited)

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.
A	US 2011/0287968 A1 (WEINHAUSEL et al) 24 November 2011 (24.11.2011) entire document	1, 3, 5-24, 26-28
A	ASHKTORAB et al. "DNA Methylation and Colorectal Cancer," Curr Colorectal Cancer Rep, 01 December 2014 (01.12.2014), Vol. 10, Iss. 4, Pgs. 425-430. entire document	1, 3, 5-24, 26-28
A	US 2016/0003829 A1 (MORPHOTEK, INC.) 07 January 2016 (07.01.2016) entire document	1, 3, 5-24, 26-28
A	WO 2015/047910 A1 (US BIOMARKERS, INC.) 02 April 2015 (02.04.2015) entire document	1, 3, 5-24, 26-28
A	GERHARDT et al. "The androgen-regulated Calcium-Activated Nucleotidase 1 (CANT1) is commonly overexpressed in prostate cancer and is tumor-biologically relevant in vitro,"The American Journal of Pathology, 22 March 2011 (22.03.2011), Vol. 178, Iss. 4, Pgs. 1847-1860. entire document	1, 3, 5-24, 26-28
P, X	HAO et al. "DNA methylation markers for diagnosis and prognosis of common cancers," Proceedings of the National Academy of Sciences of the United States of America, 26 June 2017 (26.06.2017), Vol. 114, Iss. 28, Pgs. 7414-7419 and Supplemental Information. entire document; see Supplemental Table 4	1, 3, 5-24, 26-28

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 November 2017	08 DEC 2017

Name and mailing address of the ISA/US	Authorized officer
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

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 1-28 are drawn to probes and biomarkers, methods and kits comprising the same.

The first invention of Group I+ is restricted to a probe and a biomarker, methods and kits comprising the same, wherein the probe is selected to be of the structure of Formula Ia, where A is SEQ ID NO:1, B is SEQ ID NO:2, and L is a linker of 15 nucleotides in length; the biomarker is selected to be cg08088171
(calcium activated nucleotidase 1); wherein the probe hybridizes to the biomarker. It is believed that claims 1, 3, 5-24, and 26-28 read on this first named invention and thus these claims will be searched without fee to the extent that they read on a biomarker of cg08088171 and a probe of the structure of Formula Ia, where A is SEQ ID NO:1, B is SEQ ID NO:2, and L is a linker of 15 nucleotides in length.

Applicant is invited to elect additional biomarkers and/or probes, each with specified structure and SEQ ID NO(s), to be searched in a specific combination by paying an additional fee for each set of election. An exemplary election would be a probe and a biomarker, methods and kits comprising the same, wherein the probe is selected to be of the structure of Formula Ia, where A is SEQ ID NO:3, B is SEQ ID NO:4, and L is a linker of 30 nucleotides in length; the biomarker is selected to be cg27536151 (mitogen-activated protein kinase kinase 5); wherein the probe hybridizes to the biomarker. Additional biomarkers and/or probes 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 for generating a methylation profile of a biomarker in a subject suspected of having colon cancer, requiring the selection of alternatives for the biomarker and/or probe, where "the methylation profile comprises one or more biomarkers selected from: cg08088171, cg27536151, cg13420112, cg14642259, cg20973720, cg24741563 and cg22979615" and "a methylation profile comprising one or more biomarkers selected from: cg13510262 (Sp1 transcription factor), cg13859324 (Unc-45 myosin chaperone B), and cg26952662 (collagen triple helix repeat containing 1)" and "the probe comprises at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a sequence selected from SEQ ID NOs: 1-15".

The Groups I+ share the technical features of a method of generating a methylation profile of a biomarker in a subject in need thereof, comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject; b) detecting a hybridization between the extracted genomic DNA and a probe, wherein the probe hybridizes to a biomarker; and c) generating a methylation profile based on the detected hybridization between the extracted genomic DNA and the probe; a method of selecting a subject suspected of having colon cancer for treatment, the method comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having colon cancer; b) generating a methylation profile comprising one or more biomarkers from the extracted genomic DNA; c) comparing the methylation profile of the one or more biomarkers with a control; d) identifying the subject as having colon cancer if the methylation profile correlates to the control; and e) administering an effective amount of a therapeutic agent to the subject if the subject is identified as having colon cancer; a method of determining a colon cancer subtype in a subject in need thereof, comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject; b) generating a methylation profile comprising biomarkers from the extracted genomic DNA; c) comparing the methylation profile of the biomarkers with a control; d) based on the methylation profile of the biomarkers relative to the control, identify a colon cancer subtype in the subject; and e) administering a tailored therapeutic regimen to treat the subject having the colon cancer subtype; a method of determining the prognosis of a subject having acute myeloid colon cancer or monitoring the progression of colon cancer in the subject, comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject having colon cancer; b) generating a methylation profile comprising one or more biomarkers from the extracted genomic DNA; c) obtaining a methylation score based on the methylation profile of the one or more biomarkers; and d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission, initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent; a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in the subject, comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides, wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having acute myeloid colon cancer; b) generating a methylation profile comprising one or more biomarkers from the extracted genomic DNA; c) obtaining a methylation score based on the methylation profile of the one or more biomarkers; and d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission, initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent. However, these shared technical features do not represent a contribution over the prior art.

Specifically, US 2011/0287968 A1 to Weinhausel et al. discloses a method of generating a methylation profile of a biomarker in a subject in need thereof (the present invention discloses a method of generating subsets of methylation specific markers from a set, having diagnostic power for various diseases... cancer, Abstract; a binary tree classifier for utilizing gene methylation profile can be used to predict the class of future samples. The first node of the tree incorporated a binary classifier that distinguished two subsets of the total set of classes, Para. [0040], colorectal cancer, Para. [0112]), comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides (The present invention provides a multiplexed methylation testing method which 1) outperforms the "classification" success when compared to genome-wide screenings via RNA-expression profiling, Para. [0024]; differentiate the sample type from which the methylated DNA is isolated, Para. [0026]; The methylation status can be determined by any method known in the art including methylation dependent bisulfite deamination (and consequently the identification of mC—methylated C—changes by any known methods, Para. [0047]), wherein the extracted genomic DNA is obtained from a biological sample from the subject (The set can also be optimized for a specific sample type in which the methylated DNA is tested. Such samples include blood, Para. [0112]; sample comprising DNA from a subject or patient, Para. [0179]); b) detecting a hybridization between the extracted genomic DNA and a probe (Also provided is a set of nucleic acid primers or hybridization probes being specific for a potentially methylated region of marker genes, Para. [0109]), wherein the probe hybridizes to a biomarker (detection of hybridized biotinylated PCR amplicons, Para. [0205]); and c) generating a methylation profile based on the detected hybridization between the extracted genomic DNA and the probe (invention any such samples can be used for the method of generating an inventive subset, including fixed samples, Para. [0023]; The first step of the inventive method of generating a subset, step a) of obtaining data of the methylation status, preferably comprises determining data of the methylation status, preferably by methylation specific PCR analysis, methylation specific digestion analysis. Methylation specific digestion analysis can include either or both of hybridization of suitable probes for detection to non-digested fragments or PCR amplification and detection of non-digested fragments, Para. [0033]); the method comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides (differentiate the sample type from which the methylated DNA is isolated, Para. [0026]; The methylation status can be determined by any method known in the art including methylation dependent bisulfite deamination (and consequently the identification of mC—methylated C—changes by any known methods, Para. [0047]), wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having colon cancer (The set can also be optimized for a specific sample type in which the methylated DNA is tested. Such samples include blood, Para. [0112]; sample comprising DNA from a subject or patient, Para. [0179]; Subjecting these amplicons to the methylation test, it was possible to successfully distinguish DNA from sensitive cases, e.g. distinguishing leukemia (CML) from normal healthy controls, Para. [0030]; for example breast cancer, colorectal cancer, Para. [0112]); b) generating a methylation profile comprising one or more biomarkers from the extracted genomic DNA; c) comparing the methylation profile of the one or more biomarkers with a control (invention any such samples can be used for the method of generating an inventive subset, including fixed samples, Para. [0023]; The first step of the inventive method of generating a subset, step a) of obtaining data of the methylation status, preferably comprises determining data of the methylation status, preferably by methylation specific PCR analysis, methylation specific digestion analysis. Methylation specific digestion analysis can include either or both of hybridization of suitable probes for detection to non-digested fragments or PCR amplification and detection of non-digested fragments, Para. [0033]); d) identifying the subject as having colon cancer if the methylation profile correlates to the control (markers are suitable to diagnose neoplastic disease (chronic myeloid leukemia), Para. [0120]; comparing the methylation status with the status of a confirmed disease or tumor type positive and/or negative state, thereby identifying the disease or tumor type in the sample., Para. [0046]; these markers can be used to distinguish between nodule positive conditions (malign and benign tumors) and normal controls, Para. [0157]); comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides (differentiate the sample type from which the methylated DNA is isolated, Para. [0026]; The methylation status can be determined by any method known in the art including methylation dependent bisulfite deamination (and consequently the identification of mC—methylated C—changes by any known methods, Para. [0047]), wherein the extracted genomic DNA is obtained from a biological sample from the subject; b) generating a methylation profile comprising biomarkers from the extracted genomic DNA (The set can also be optimized for a specific sample type in which the methylated DNA is tested. Such samples include blood, Para. [0112]; sample comprising DNA from a subject or patient, Para. [0179]; Subjecting these amplicons to the methylation test, it was possible to successfully distinguish DNA from sensitive cases, e.g. distinguishing leukemia (CML) from normal healthy controls, Para. [0030]; for example breast cancer, colorectal cancer, Para. [0112]); c) comparing the methylation profile of the biomarkers with a control (invention any such samples can be used for the method of generating an inventive subset, including fixed samples, Para. [0023]; The first step of the inventive method of generating a subset, step a) of obtaining data of the methylation status, preferably comprises determining data of the methylation status, preferably by methylation specific PCR analysis, methylation specific digestion analysis. Methylation specific digestion analysis can include either or both of hybridization of suitable probes for detection to non-digested fragments or PCR amplification and detection of non-digested fragments, Para. [0033]); d) based on the methylation profile of the biomarkers relative to the control (invention any such samples can be used for the method of generating an inventive subset, including fixed samples, Para. [0023]; The first step of the inventive method of generating a subset, step a) of obtaining data of the methylation status, preferably comprises determining data of the methylation status, preferably by methylation specific PCR analysis, methylation specific digestion analysis. Methylation specific digestion analysis can include either or both of hybridization of suitable probes for detection to non-digested fragments or PCR amplification and detection of non-digested fragments, Para. [0033]), comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides (differentiate the sample type from which the methylated DNA is isolated, Para. [0026]; The methylation status can be determined by any method known in the art including methylation dependent bisulfite deamination (and consequently the identification of mC—methylated C—changes by any known methods, Para. [0047]), wherein the extracted genomic DNA is obtained from a biological sample from the subject having colon cancer (The set can also be optimized for a specific sample type in which the methylated DNA is tested. Such samples include blood, Para. [0112]; sample comprising DNA from a subject or patient, Para. [0179]; Subjecting these amplicons to the methylation test, it was possible to successfully distinguish DNA from sensitive cases, e.g. distinguishing leukemia (CML) from normal healthy controls, Para. [0030];]; for example breast cancer, colorectal cancer, Para. [0112]); b) generating a methylation profile comprising one or more biomarkers from the extracted genomic DNA; c) obtaining a methylation score based on the methylation profile of the one or more biomarkers (invention any such samples can be used for the method of generating an inventive subset, including fixed samples, Para. [0023]; The first step of the inventive method of generating a subset, step a) of obtaining data of the methylation status, preferably comprises determining data of the methylation status, preferably by methylation specific PCR analysis, methylation specific digestion analysis. Methylation specific digestion analysis can include either or both of hybridization of suitable probes for detection to non-digested fragments or PCR amplification and detection of non-digested fragments, Para. [0033]) comprising: a) processing an extracted genomic DNA with a deaminating agent to generate a genomic DNA sample comprising deaminated nucleotides (differentiate the sample type from which the methylated DNA is isolated, Para. [0026]; The methylation status can be determined by any method known in the art including methylation dependent bisulfite deamination (and consequently the identification of mC—methylated C—changes by any known methods, Para. [0047]), wherein the extracted genomic DNA is obtained from a biological sample from the subject suspected of having colon cancer (these markers are suitable to diagnose neoplastic disease (chronic myeloid leukemia), Para. [0120]; The set can also be optimized for a specific sample type in which the methylated DNA is tested. Such samples include blood, Para. [0112]; sample comprising DNA from a subject or patient, Para. [0179]; Subjecting these amplicons to the methylation test, it was possible to successfully distinguish DNA from sensitive cases, e.g. distinguishing leukemia (CML) from normal healthy controls, Para. [0030]); b) generating a methylation profile comprising one or more biomarkers from the extracted genomic DNA (invention any such samples can be used for the method of generating an inventive subset, including fixed samples, Para. [0023]; The first step of the inventive method of generating a subset, step a) of obtaining data of the methylation status, preferably comprises determining data of the methylation status, preferably by methylation specific PCR analysis, methylation specific digestion analysis. Methylation specific digestion analysis can include either or both of hybridization of suitable probes for detection to non-digested fragments or PCR amplification and detection of non-digested fragments, Para. [0033]).

Further, US 2016/0003829 A1 to Morphotek, Inc. discloses a method of selecting a subject suspected of having colon cancer for treatment (Provided herein are methods for determining the risk that a subject diagnosed with colorectal cancer will develop a recurrence of colorectal cancer and methods of predicting clinical outcome, Abstract;), and e) administering an effective amount of a therapeutic agent to the subject if the subject is identified as having colon cancer (if a subject is determined to be at high risk for recurrence of colorectal cancer, further therapy may be elected or administered, Para. [0062]); a method of determining a colon cancer subtype in a subject in need thereof; identify a colon cancer subtype in the subject (the methods for determining the risk that a subject diagnosed with colorectal cancer will develop a recurrence of colorectal cancer involve a) determining the level of expression for each marker of a panel of markers in a panel of tumor compartments in a tumor tissue sample from the subject,, Para. [0063]); and e) administering a tailored therapeutic regimen to treat the subject having the colon cancer subtype (); a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in the subject (information is useful to the patient and the physician for assessing the risk versus benefit of observation versus adjuvant therapy for that patient. For example, if a subject is determined to be at high risk for recurrence of colorectal cancer, further therapy may be elected or administered. Such therapy may include TEM-1-targeted therapy (e.g., MORAb-004), chemotherapy, radiation therapy, and/or monitoring for disease recurrence or progression, Para. [0062]); a biological sample from the subject having colon cancer (s in a tumor tissue sample from the subject, Para. [0062]); and d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission (determining the subject's TAPPS score; and c) comparing the TAPPS score of the subject to the TAPPS score of a population of subjects diagnosed with colorectal cancer, Para. [0063]; patients most likely to benefit from adjuvant therapy or from a more aggressive monitoring of disease recurrence can be identified. Conversely, patients who may be candidates for adjuvant therapy based on current prognostic criteria, but who are identified as being at low risk for recurrence based on this assay, may avoid the unnecessary risks associated with existing adjuvant therapy, Para. [0066]), initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent (if a subject is determined to be at high risk for recurrence of colorectal cancer, further therapy may be elected or administered, Para. [0062]); a method of determining the prognosis of a subject having colon cancer or monitoring the progression of colon cancer in the subject (information is useful to the patient and the physician for assessing the risk versus benefit of observation versus adjuvant therapy for that patient. For example, if a subject is determined to be at high risk for recurrence of colorectal cancer, further therapy may be elected or administered. Such therapy may include TEM-1-targeted therapy (e.g., MORAb-004), chemotherapy, radiation therapy, and/or monitoring for disease recurrence or progression, Para. [0062]); obtaining a methylation score based on the methylation profile of the one or more biomarkers; and d) based on the methylation score, initiate a first treatment, decrease a dosage of a first therapeutic agent if the subject has experienced a remission, initiate a second treatment if the subject has experienced a relapse, or switch to a second therapeutic agent if the subject becomes refractory to the first therapeutic agent (information is useful to the patient and the physician for assessing the risk versus benefit of observation versus adjuvant therapy for that patient. For example, if a subject is determined to be at high risk for recurrence of colorectal cancer, further therapy may be elected or administered. Such therapy may include TEM-1-targeted therapy (e.g., MORAb-004), chemotherapy, radiation therapy, and/or monitoring for disease recurrence or progression, Para. [0062]).

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