

DNA Methylation Markers for Noninvasive Detection of Cancer and Uses Thereof

5

FIELD OF THE INVENTION

The present invention relates to DNA methylation signatures in human DNA, particularly in the field of molecular diagnostics.

10

BACKGROUND OF THE INVENTION

Cancer has become a major killer of humans. Early detection of cancer can significantly improve cure rates and reduce the horrific personal and financial cost to the patients their families and the health care system. For example, Hepatocellular Carcinoma (HCC) is the fifth most common cancer world-wide (El-Serag, 2011). It is particularly prevalent in Asia, and its occurrence is highest in areas where hepatitis B is prevalent, indicating a possible causal relationship (Flores & Marrero, 2014). Follow up of high-risk populations such as chronic hepatitis patients and early diagnosis of transitions from chronic hepatitis to HCC would improve cure rates. The survival rate of hepatocellular carcinoma is currently extremely low because it is almost always diagnosed at the late stages. Liver cancer could be effectively treated with cure rates of >80% if diagnosed early. Advances in imaging have improved noninvasive detection of HCC (Tan, Low, & Thng, 2011; Valente et al., 2014). However, current diagnostic methods, which include imaging and immunoassays with single proteins such as alpha-fetoprotein often fail to diagnose HCC early (Flores & Marrero, 2014). These challenges are not limited to HCC but common to other cancers as well. For example, early detection of breast cancer and colorectal cancer could dramatically reduce morbidity and mortality and the cost to the public health system and insurance companies. Moreover, certain cancers such as pancreatic cancers are detected almost invariably late resulting in virtually certain mortality. Advances in imaging have improved early detection of

cancers, however high-resolution imaging such as MRI is expensive, requires highly trained personnel and is unavailable in many locations. It has not evolved yet to a method of screening of wide populations. To have an impact on reducing morbidity and mortality from cancer it is necessary to develop a noninvasive, robust but nevertheless low-cost
5 method that could be used in wide geographic areas for routine screening of the population. The main challenge is that solid tumors hide in internal organs and evolve long before they exhibit clinical symptoms. It is however possible to obtain tumor material noninvasively.

It is widely established by now that tumor DNA is shed into the system and could
10 be found in plasma (Warton & Samimi, 2015) and possibly other secreted body fluids such as urine and saliva, as well as feces. By measuring molecular characteristics of tumor DNA, it is possible to determine that the DNA found in body fluids originated in the tumor (Zhai et al., 2012). Although tumor cells develop mutations that could distinguish tumor DNA from normal cells DNA, the number of possible mutations is vast
15 and common mutations don't occur in all tumors (Dominguez-Vigil, Moreno-Martinez, Wang, Roehrl, & Barrera-Saldana, 2018).

DNA methylation, a covalent modification of DNA, which is a primary mechanism of epigenetic regulation of genome function is ubiquitously altered in tumors (Aguirre-Ghiso, 2007; Baylin et al., 2001; Ehrlich, 2002; Issa et al., 1993). DNA
20 methylation profiles of tumors are potentially robust tools for tumor classification, prognosis and prediction of response to chemotherapy (Stefansson et al., 2014). The major drawback for using tumor DNA methylation in early diagnosis is that it requires invasive procedures and anatomical visualization of the suspected tumor. Circulating tumor cells are a noninvasive source of tumor DNA and are used for measuring DNA
25 methylation in tumor suppressor genes (Radpour et al., 2011). Hypomethylation of HCC DNA is detectable in patients' blood (Ramzy, Omran, Hamad, Shaker, & Abboud, 2011) and genome wide bisulfite sequencing was recently applied to detect hypomethylated DNA in plasma from HCC patients (Chan et al., 2013). However, this source is limited, particularly at early stages of cancer and the DNA methylation profiles are confounded
30 by host DNA methylation profiles. Genome wide bisulfite sequencing is a relatively costly procedure and requires significant bioinformatics analysis which makes it

unfeasible as a screening tool. The challenge is therefore to delineate a small number of CGs that could robustly differentiate tumor DNA from nontumor DNA and develop a high throughput low cost assay that will enable the screening of wide populations in broad and diverse geographic areas. More recently several groups have performed comparative analysis of genome wide DNA methylation maps of cancer and normal DNA and blood DNA (Zhai et al., 2012). However, the main challenge with these approaches is that they have not taken into account cell free DNA from other tissues that is found in blood at different levels that are unanticipated a priori. Contaminating DNA from another tissue that has a similar methylation profile as the cancer tissue could lead to false positives. In addition, past approaches have quantitatively compared DNA methylation in normal and cancer tissues. This quantitative difference is diluted when tumor DNA is mixed with different and unknown amounts of DNA from other untransformed tissues, which can cause false negatives. These deficiencies in current methods necessitate a different approach that is disclosed in the present inventive subject matter.

Further publications dealing with ore relating to the use of systems and methods for detecting cancer are: Grigg G, Clark S. Sequencing 5-methylcytosine residues in genomic DNA. *Bioessays*. 1994 June; 16(6):431-6, 431; Zeschnigk M, Schmitz B, Dittrich B, Buiting K, Horsthemke B, Doerfler W. Imprinted segments in the human genome: different DNA methylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. *Hum Mol Genet*. 1997 March; 6(3):387-95; Feil R, Charlton J, Bird A P, Walter J, Reik W. Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res*. 1994 February 25;22(4):695-6; Martin V, Ribieras S, Song-Wang X, Rio M C, Dante R. Genomic sequencing indicates a correlation between DNA hypomethylation in the 5' region of the pS2 gene and its expression in human breast cancer cell lines. *Gene*. 1995 May 19;157(1-2):261-4; WO 97 46705, WO 95 15373 and WO 45560. Aguirre-Ghiso, J. A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. *Nat Rev Cancer*, 7(11), 834-846. doi:10.1038/nrc2256

- Baylin, S. B., Esteller, M., Rountree, M. R., Bachman, K. E., Schuebel, K., & Herman, J. G. (2001). Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet*, 10(7), 687-692.
- Breitbart, S., Tug, S., Helmig, S., Zahn, D., Kubiak, T., Michal, M., . . . Simon, P. (2014). Direct quantification of cell-free, circulating DNA from unpurified plasma. *PLoS One*, 9(3), e87838. doi:10.1371/journal.pone.0087838
- Chan, K. C., Jiang, P., Chan, C. W., Sun, K., Wong, J., Hui, E. P., . . . Lo, Y. M. (2013). Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci U S A*, 110(47), 18761-18768. doi:10.1073/pnas.1313995110
- Dominguez-Vigil, I. G., Moreno-Martinez, A. K., Wang, J. Y., Roehrl, M. H. A., & Barrera-Saldana, H. A. (2018). The dawn of the liquid biopsy in the fight against cancer. *Oncotarget*, 9(2), 2912-2922. doi:10.18632/oncotarget.23131
- Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene*, 21(35), 5400-5413.
- El-Serag, H. B. (2011). Hepatocellular carcinoma. *N Engl J Med*, 365(12), 1118-1127. doi:10.1056/NEJMr1001683
- Flores, A., & Marrero, J. A. (2014). Emerging trends in hepatocellular carcinoma: focus on diagnosis and therapeutics. *Clin Med Insights Oncol*, 8, 71-76. doi:10.4137/CMO.S9926
- Issa, J. P., Vertino, P. M., Wu, J., Sazawal, S., Celano, P., Nelkin, B. D., . . . Baylin, S. B. (1993). Increased cytosine DNA-methyltransferase activity during colon cancer progression. *J Natl Cancer Inst*, 85(15), 1235-1240.
- Luczak, M. W., & Jagodzinski, P. P. (2006). The role of DNA methylation in cancer development. *Folia Histochem Cytobiol*, 44(3), 143-154.
- Radpour, R., Barekati, Z., Kohler, C., Lv, Q., Burki, N., Diesch, C., . . . Zhong, X. Y. (2011). Hypermethylation of tumor suppressor genes involved in critical regulatory pathways for developing a blood-based test in breast cancer. *PLoS One*, 6(1), e16080. doi:10.1371/journal.pone.0016080

- Ramzy, II, Omran, D. A., Hamad, O., Shaker, O., & Abboud, A. (2011). Evaluation of serum LINE-1 hypomethylation as a prognostic marker for hepatocellular carcinoma. *Arab J Gastroenterol*, 12(3), 139-142. doi:10.1016/j.ajg.2011.07.002
- Stefansson, O. A., Moran, S., Gomez, A., Sayols, S., Arribas-Jorba, C., Sandoval, J., . . . Esteller, M. (2014). A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol*. doi:10.1016/j.molonc.2014.10.012
- Tan, C. H., Low, S. C., & Thng, C. H. (2011). APASL and AASLD Consensus Guidelines on Imaging Diagnosis of Hepatocellular Carcinoma: A Review. *Int J Hepatol*, 2011, 519783. doi:10.4061/2011/519783
- Valente, S., Liu, Y., Schneckeburger, M., Zwergel, C., Cosconati, S., Gros, C., . . . Mai, A. (2014). Selective non-nucleoside inhibitors of human DNA methyltransferases active in cancer including in cancer stem cells. *J Med Chem*, 57(3), 701-713. doi:10.1021/jm4012627
- Warton, K., & Samimi, G. (2015). Methylation of cell-free circulating DNA in the diagnosis of cancer. *Front Mol Biosci*, 2, 13. doi:10.3389/fmolb.2015.00013
- Xu, R. H., Wei, W., Krawczyk, M., Wang, W., Luo, H., Flagg, K., . . . Zhang, K. (2017). Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater*, 16(11), 1155-1161. doi:10.1038/nmat4997
- Zhai, R., Zhao, Y., Su, L., Cassidy, L., Liu, G., & Christiani, D. C. (2012). Genome-wide DNA methylation profiling of cell-free serum DNA in esophageal adenocarcinoma and Barrett esophagus. *Neoplasia*, 14(1), 29-33.

SUMMARY

Embodiments of the claimed subject matter show that cancer is associated with a set of “categorically” distinct DNA methylation signatures that are different from any normal tissue and blood cell DNA methylation profiles. These sites create a binary differentiation between cancer and other tissues, whereby these sites are only methylated in cancer and fully unmethylated in other cancers. Thus, it is possible using deep next generation sequencing to detect even a few molecules of cancer cells on the background

of normal cell DNA profile of methylation. Embodiments of the inventive subject matter enable detection of cell free tumor DNA even on a high background of cell free DNA from other tissues and are thus particularly suitable for early detection of cancer using cell free (CF) DNA extracted from body fluids, for example saliva, plasma, urine, feces etc. Embodiments also allow for early detection of cancer in tissue smears such as pap smears as well as biopsies and needle biopsies. Previous analyses in the prior art only compared normal and cancer cells from the same tissue and blood and derived sites that are quantitatively different in their DNA methylation level (Xu et al., 2017). However, sites discovered by such prior art analyses can't detect CF tumor DNA when it is mixed with other tissue CF DNA (see FIG. 2 for ctDNA markers for HCC from Sun Yat Sen University Cancer hospital). One embodiment of the present claimed subject matter reveals a unique set of sites that are unmethylated in all tissues but methylated in specific cancers. Another embodiment reveals a method to discover categorically distinct methylation sites in cancers, other tissues and other diseases called the "binary-categorical differentiation (BCD) method" using different sources of genome wide DNA methylation data derived by next generation sequencing, MeDIP arrays, MeDIP sequencing, and the like. One embodiment reveals a combination of "Categorical" DNA methylation sites for detection of a. hepatocellular carcinoma (HCC), b. lung cancer, c. prostate cancer, d. breast cancer, e. colorectal cancer, f. head and neck squamous cell carcinoma (HNSC). g. pancreatic cancer, h. brain cancer (glioblastoma), i. gastric cancer j. ovarian cancer, k. cervical cancer, l. esophageal carcinoma m. bladder cancer, n. renal cancer, o. testicular cancer, p. common solid tumors, q. blood cancer profiles in a discovery set of genome wide data. Another embodiment also reveals a combination of "Categorical" DNA methylation sites that differentiate tumors by their tissue of origin. This embodiment differentiates the assay from prior art methods for detecting methylated CF DNA which have low tissue specificity. Embodiments validate the polygenic DNA methylation assays for detection of cancer in DNA methylation data from hundreds of patients as well as the tissue of origins of the tumors with high sensitivity and specificity. The present invention discloses a method that accurately measures DNA methylation in a polygenic set of CG IDs in hundreds of people concurrently, by sequential amplification with target specific primers followed by barcoding primers and multiplexed sequencing

in a single next generation Miseq sequencing reaction, data extraction and quantification of methylation from a small volume of body fluids such as plasma, saliva or urine.

Another embodiment of the inventive subject matter also discloses measurement of methylation of the said DNA methylation CG IDs using pyrosequencing assays or methylation specific PCR. Another embodiment discloses the calculation of either a “categorical” or a polygenic weighted methylation score that differentiates people with cancer from healthy people. Another embodiment discloses a novel process leading from plasma, urine, feces, tissue biopsy or tissue swabs to prediction of cancer in a person with no other clinical evidence for cancer. Another embodiment could be used by any person skilled in the art to detect cancer as well as other diseases that involve cell death and release of CF DNA such Alzheimer disease and other neurodegenerative diseases for neurons, heart disease for heart muscle cells. The DNA methylation markers (CG IDs) described in the embodiments will be utilized for a. noninvasive early detection of cancer in otherwise “healthy” people through routine “checkup”; b. monitoring “high risk” people such as chronic hepatitis patients who are at high risk for HCC or smokers who are at high risk for lung cancer; c. monitoring response to therapy in patients undergoing cancer treatment to detect recurrence or metastasis.

Embodiments demonstrate the utility detecting cancer of unknown samples using polygenic or categorical scores based on the DNA methylation measurement methods disclosed herein. The disclosed embodiments could be used by any person skilled in the art to detect cancer in body fluids, feces, urine and tissues of any cancer or diseased tissue using any method for methylation analysis that are available to those skilled in the art such as for example next generation bisulfite sequencing, Illumina Epic microarrays, capture sequencing, methylated DNA Immunoprecipitation (MeDIP) methylation specific PCR and any methylation measurements that becomes available.

Embodiments also disclose the potential for discovery of new “polygenic” categorical DNA methylation markers for other cancers and diseases using any method available to people skilled in the art for genome wide sequencing such as next generation bisulfite sequencing, MeDip sequencing, ion torrent sequencing, Epic microarrays etc. followed by binary-categorical differentiation (BCD) method of analysis for discovering specific and sensitive markers that will be used for noninvasive detection of disease

Embodiments of the present inventive subject matter include:

In a first aspect, embodiments provide polygenic DNA methylation markers of cancer in cell free DNA in body fluids such as plasma for early detection of cancer, said polygenic DNA methylation markers set being derived using “binary-categorical differentiation (BCD) analysis” as disclosed herein on genome wide DNA methylation derived by mapping methods such as Illumina 450K or EPIC arrays, genome wide bisulfite sequencing, methylated DNA Immunoprecipitation (MeDIP) sequencing or hybridization with oligonucleotide arrays.

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below (or a short subset of this list such as the example listed below Table 1) for early detection of Liver cancer hepatocellular carcinoma (HCC) using plasma CF DNA or other body fluid CF DNA.

Table 1.

Liver_detect					
cg00370303	cg10900437	cg02012576	cg16460359	cg04035559	cg17419241
cg00931619	cg11223367	cg03768777	cg16977570	cg04085025	cg18607529
cg05040544	cg19289599	cg06233293	cg24804544	cg26523670	cg09992116
cg05739190	cg24599205				

Subset for detect:

cg02012576, cg03768777, cg24804544, cg05739190

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example shown below Table 2) for specifying the origin of the cancer as HCC and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

Table 2.

Liver_spec					
cg12137206	cg14126493	cg06105778	cg22076972	cg13341720	cg03705926
cg09363194	cg07036412	cg02702614	cg10181419	cg05876864	cg11068343
cg17167468	cg15375239	cg00026222	cg17283781	cg16147221	cg26386472
cg14570307	cg06207432	cg07610192	cg03422204	cg11684022	cg23693289
cg21107197	cg04920951	cg20385508	cg25296314	cg20707679	cg26703661
cg00456086	cg05009389	cg19388016	cg08460435	cg04739306	cg04221886
cg26797073	cg04109768	cg05337743	cg00483503	cg18668780	cg10604002
cg27650175	cg05684891	cg26026416	cg00177496	cg14221460	cg16551483
cg13438961	cg24432073	cg21059834	cg23305567	cg04809136	cg21105227

Subset for specificity:

cg14126493

- In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 3) for early detection of lung cancer using plasma CF DNA or other body fluid CF DNA.

Table 3.

Lung_detect					
cg04223424	cg06530490	cg08944430	cg09463882	cg11017065	cg12405785
cg16405026	cg21410080	cg23141355	cg25024074		

10

Subset for detect:

cg04223424, cg23141355

- In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 4) for specifying the origin of the cancer as lung cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

20

Table 4

Lung_spec					
cg05917732	cg02682457	cg23407396	cg23601095	cg23141355	cg15386964
cg02578368	cg24631970	cg27487839	cg16405026		
cg23141355	cg06530490	cg04223424	cg25470077	cg07138603	cg23460835
cg20678442	cg15436096				

5

Subset for spec:

cg05917732, cg25470077

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below (or a short subset of this list such as the example listed below Table 5) for early detection of prostate cancer as well as for specifying the origin of the cancer as prostate cancer and discriminating from other 16 common solid tumor cancers using plasma CF DNA or other body fluid CF DNA.

15 **Table 5.**

Prostate_detect_spec			
cg14283569	cg18085998	cg13303553	cg17929627

Subset for detect_spec:

cg14283569

[it is a subset of the 4 listed in the table above]

20

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 6) for early detection of breast cancer using plasma CF DNA or other body fluid CF DNA.

5

Table 6.

Breast_detect					
cg13031251	cg19595750	cg16842053	cg18565473	cg09734791	cg25694349
cg09695735	cg03637878	cg26261793	cg18800085	cg16132520	cg05617413
cg17228900	cg06945936	cg08406370	cg24427504	cg26937500	cg11297107
cg02215070	cg14140647	cg05377226	cg07070305	cg24899571	cg07844931

Subset for detect:

10 cg13031251, cg09734791, cg09695735, cg03637878

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 7) for specifying the origin of the cancer as breast cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

15

Table 7.

Breast_Spec				
cg00467244	cg04194674	cg06879394	cg10720997	cg19966212
cg00722320	cg04751811	cg06998282	cg11498607	cg24525457
cg01308827	cg06282270	cg08066035	cg14862207	cg26228266
cg03113878	cg06405186	cg08296680	cg17945153	cg20180843

20 Subset for spec:

cg03113878, cg20180843

- In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 8) for early detection of colorectal cancer (CRC) as well as for specifying the origin of the cancer as colorectal cancer and discriminating from other 16 common solid tumor cancers using plasma CF DNA or other body fluid CF DNA.

Table 8

CRC_detect_spec					
cg08808128	cg03788131	cg09854653	cg21627760	cg08169901	cg07494047
cg01566242	cg13788592	cg24102266	cg17716617	cg16733654	

Subset for detect-spec

- 10 cg09854653, cg01566242

- In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 9) for early detection of pancreatic cancer using plasma CF DNA or other body fluid CF DNA.

Table 9.

Pancreas_detect					
cg11017065	cg23833588	cg25024074	cg21277995	cg22286978	cg06241792
cg11591516	cg10096177	cg12035092	cg03611007	cg17996329	cg13807970
cg16678602	cg15386964	cg01423964	cg07900968	cg19118812	cg06728579
cg16232979	cg08406370				

- 20 Subset for detect:
cg25024074, cg15386964, cg16232979

- In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 10) for specifying the origin of the cancer as pancreatic

cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

Table 10

Pancreas_Spec					
cg01237565	cg08182975	cg15323936	cg19102272	cg01311909	cg26927232
cg22290704	cg15832577	cg20983577	cg26466027	cg09796911	cg15850155
cg15510118	cg25591377	cg16165258	cg25595541	cg18151519	cg19749445
cg14870128	cg13441142				

Subset for spec:

cg01237565, cg08182975, cg20983577, cg25591377

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 11) for early detection of brain cancer (glioblastoma) as well as for specifying the origin of the cancer as brain cancer (glioblastoma) and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid CF DNA.

Table 11

Brain_detect-spec					
cg06887260	cg19116006	cg10938374	cg03663746	cg25568243	cg27449131
cg24917627	cg04134305	cg09797645	cg02892595	cg13231951	cg26269703
cg09183941	cg16842053	cg26551897	cg26988692	cg07849581	cg25026703
cg06798642	cg22963915	cg04245373	cg25533556	cg24917627	cg27659841
cg04692993	cg06045408	cg06887260	cg11345323	cg17167076	cg17526812
cg19929355	cg22513169	cg22865720			

Subset for spec-detect

Cg19929355

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 12) for early detection of stomach (gastric) cancer using plasma CF DNA or other body fluid CF DNA.

5

Table 12.

Stomach_detect					
cg04125371	cg05377226	cg05611779	cg06241792	cg07900968	cg09734791
cg11017065	cg12510981	cg13807970	cg15760257	cg18323466	cg19118812
cg19419279	cg19769760	cg20334243	cg26261793		

Subset for detect:

cg05611779, cg09734791, cg15760257

10

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example shown below Table 13) for specifying the origin of the cancer as gastric cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

15

Table 13.

Stomach_Spec					
cg17187167	cg07350470	cg05000488	cg25612145	cg09851120	cg00904726
cg04911739	cg18192294	cg11861709	cg19452853	cg02706110	cg03768513
cg05611779	cg06981182	cg06118999	cg04812509	cg10131095	cg05339066
cg19235339					

Subset for spec:

20 cg05611779, cg19235339

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example shown below Table 14) for early detection of ovarian cancer using plasma CF DNA or other body fluid CF DNA.

25

Table 14.

Ovarian_detect					
~~cg24339193~~	cg04008429	cg06849719	~~cg22694153~~	~~cg11252337~~	cg15804105
cg12479674	cg06961071	~~cg21210985~~	cg01556502	cg02537149	cg23983315
cg03597143	cg27209395				

Subset for detect:

cg24339193, cg22694153, cg11252337, cg21210985

5

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 15) for specifying the origin of the cancer as ovarian cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

10

Table 15.

Ovarian_spec					
cg00895834	cg01159194	cg01961086	cg02649698	cg03345116	cg03456771
cg09392827	cg10178270	cg10581012	cg13459217	cg15701612	cg17130982
cg05901462	~~cg07068768~~	cg08389588	cg09173621	cg19276014	~~cg19846609~~
cg18476766	cg19129687				

Subset for spec:

15 cg07068768, cg19846609

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example shown below Table 16) for early detection of cervical cancer using plasma CF DNA or other body fluid CF DNA.

20

Table 16.

Cervix_detect					
cg00578154	cg13644629	cg01423964	cg15745619	cg04522671	cg21621906
cg00757182	cg14962363	cg01601746	cg17228900	cg07126167	cg22806837

cg08134106	cg22922289	cg09260640	cg24625128	cg11259628	cg27420520
cg08535260	cg23141355	cg09734791	cg25024074		

Subset for detect:

cg00757182, cg01601746

- 5 In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 17) for specifying the origin of the cancer as cervical cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

10

Table 17.

Cervix_spec					
cg00829990	cg02401399	cg04996219	cg07066594	cg07195011	cg07576142
cg09260640	cg12961842	cg13668618	cg18543270		

Subset for spec:

cg07066594, cg09260640, cg12961842

15

- In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 18) for early detection of head and neck squamous cell carcinoma (HNSC) carcinoma using plasma CF DNA or other body fluid CF DNA.

20

Table 18.

HNSC_detect					
cg01613638	cg02776314	cg03280624	cg04524120	cg05151803	cg07746323
cg07900968	cg08406370	cg11108676	cg12083965	cg15397448	cg17428324
cg18403606	cg20334243	cg26770917	cg27009208	cg27420520	

Subset for detect:

cg07900968, cg20334243, cg27420520

25

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 19) for specifying the origin of the cancer as head and neck squamous cell carcinoma (HNSC) and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

Table 19.

HNSC_spec					
cg01217080	cg03522799	cg06672120	cg09136346	cg10155875	cg18006328
cg18443253	cg19287220				

Subset for spec:

10 cg18006328, cg19287220

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or in a short subset of this list (such as the example listed below Table 20) for early detection of esophageal carcinoma using plasma CF DNA or other body fluid CF DNA.

Table 20.

Esophagus_detect					
cg03280624	cg03735888	cg06530490	cg06963053	cg08944430	cg09734791
cg11017065	cg12035092	cg18344922	cg19118812	cg20334243	cg22128431
cg23141355	cg24740531	cg27009208	cg27420520		

Subset for detect:

20 cg03280624, cg03735888, cg09734791, cg27420520

In one embodiment, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 21) for specifying the origin of the cancer as esophageal carcinoma and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

Table 21.

Esophagus_spec					
cg00532449	cg02763101	cg04743654	cg08055087	cg08932440	cg09556952
cg10608333	cg12473285	cg12966367	cg17579667	cg17949440	cg18723937
cg21554552	cg22647407	cg23286646	cg23730575	cg27569446	

Subset for spec:

Cg09556952, cg12473285

5

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 22) for early detection of bladder cancer using plasma CF DNA or other body fluid CF DNA.

10

Table 22.

Bladder_detect					
cg01423964	~~cg04223424~~	cg01556502	cg23141355	~~cg10723962~~	cg09734791
~~cg25024074~~	cg21039778	cg19113641	cg04342821	cg09008417	cg05873285

Subset for detect:

cg04223424, cg10723962, cg25024074

15

In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (as shown in the example listed below Table 23) for specifying the origin of the cancer as bladder cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid DNA.

20

Table 23.

Bladder_spec					
cg14773260	cg08827322	cg05085809	cg13544006	cg02293936	cg10153335
cg05039004	cg14718568	cg07790170	cg12995090	cg16495265	cg01508045
cg04933208	cg03231163	cg06708937	cg00675569	cg06312813	cg14126688
cg12984729	cg02384661	cg08307030	cg26884027	cg20540209	cg04355159
cg08857479	cg23395715	cg22006640	cg26279336	cg19898108	cg17446010
cg12911122					

Subset for spec:

cg13544006

- 5 In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as the example listed below Table 24) for early detection of renal (kidney) cancer for specifying the origin of the cancer as renal cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid CF DNA.

10

Table 24.

Kidney_detect_spec					
cg08884571	cg00011225	cg14535274	cg05507490	cg10367244	cg07190763
cg02971546	cg19990785	cg03084949	cg08765317	cg02820958	cg23946709
cg26642667					

Subset for detect spec:

cg08884571, cg00011225, cg00011225

15

- 20 In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list (such as shown in the example listed below Table 25) for early detection of testicular cancer for specifying the origin of the cancer as testicular cancer and discriminating from other 10 common solid tumor cancers using plasma CF DNA or other body fluid CF DNA.

Table 25.

Testicular_detect_spec					
cg14531093	cg11777290	cg13283877	cg03966099	cg26939375	cg26297530
cg16157016	cg14002772	cg09890687	cg05134918	cg09719850	cg06456125
cg17978367	cg20100671	cg26627956	cg19789755	cg14393609	cg21634331
cg02039634	cg25159927	cg01895439			

Subset for detect and spec:

cg14531093, cg25159927

- 5 In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs in the list below or a short subset of this list such (as shown in the example listed below Table 26) for early detection of one of 13 most common solid tumors using plasma CF DNA or other body fluid CF DNA.

10 **Table 26.**

Pan-cancer_detect					
cg01423964	cg04223424	cg06530490	cg06543087	cg07900968	cg08406370
cg08848774	cg09734791	cg10723962	cg11017065	cg15759056	cg15760257
cg16405026	cg17228900	cg21277995	cg22286978	cg23141355	cg24427504
cg25024074	cg27420520				

Subset for detect:

cg10723962, cg15759056, cg24427504, cg25024074

- 15 In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs delineated by the BCD method on genome wide DNA methylation data as shown in Table 27 (or a short subset of this combination as shown below Table 27) for early detection of blood cancers such as AML, CLL, and the like using white blood cells, plasma CF DNA or other body fluid CF DNA.

20

Table 27

AML_detect-spec					
-----------------	--	--	--	--	--

cg00594866	cg05608159	cg18658397	cg18780412	cg20439288	cg22828045
cg25375340					

Subset for detect-spec:

cg18658397, cg18780412, cg20439288, cg22828045, cg25375340

- 5 In other embodiments, the polygenic DNA methylation markers are a combination of CG IDs shown in the list below (or a short subset of this list shown in the example listed below Table 28) for early detection of Melanoma for specifying the origin of the cancer as melanoma and discriminating from other 16 common solid tumor cancers using plasma CF DNA or other body fluid CF DNA.

10

Table 28

Melanoma_detect-spec					
cg00325866	cg01228636	cg04616691	cg04824711	cg05569109	cg06019853
cg08611163	cg10830649	cg12593303	cg15307891	cg18866529	cg19530886
cg19634213	cg20146030	cg21755725	cg22280705	cg24217704	cg24678095
cg26127345	cg27084903				

Subset for detect-spec:

cg15307891, cg18866529, cg27084903

- 15 In another aspect of the inventive subject matter, there is provided a kit and a process for detecting cancer, comprising means and reagents for detecting DNA methylation measurements of polygenic DNA methylation markers.

- 20 In one embodiment, a kit is provided for detecting hepatocellular carcinoma comprising means and reagents for DNA methylation measurements of the CG IDs of table 1 and 2.

In another embodiment, a kit is provided for detecting lung cancer comprising means and reagents for DNA methylation measurements of the CG IDs of table 3 and 4.

In another embodiment, a kit is provided for detecting prostate cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table

- 25 5.

In another embodiment, a kit is provided for detecting breast cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 6 and 7.

5 In another embodiment, a kit is provided for detecting colorectal cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 8.

In another embodiment, a kit is provided for detecting pancreatic cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 9 and 10.

10 In yet another embodiment, a kit is provided for detecting brain cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 11.

In another embodiment, a kit is provided for detecting gastric cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 15 12 and 13.

In another embodiment, a kit is provided for detecting ovarian cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 14 and 15.

20 In another embodiment, a kit is provided for detecting cervical cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 16 and 17.

In another embodiment, a kit is provided for detecting head and neck squamous carcinoma (HNSC) comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 18 and 19.

25 In another embodiment, a kit is provided for detecting esophageal carcinoma comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 20 and 21.

In another embodiment, a kit is provided for detecting bladder cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 30 22 and 23.

In another embodiment, a kit is provided for detecting renal cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 24.

5 In another embodiment, a kit is provided for detecting testicular cancer comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 25

In other embodiments, a kit is provided for detecting one of 13 common cancers (bladder, brain, breast, cervical, colorectal, esophageal, HNSC, HCC (liver), lung, ovarian, pancreatic, prostate, stomach) comprising means and reagents for detecting
10 DNA methylation measurements of the CG IDs of table 26.

In another embodiment, a kit is provided for detecting blood cancers such as AML and CLL comprising means and reagents for detecting DNA methylation measurements of the CG IDs detected by the BCD method that are specific for different
15 subtypes of blood cancer Table 27

In another embodiment, a kit is provided for detecting melanoma comprising means and reagents for detecting DNA methylation measurements of the CG IDs of table 28.
20

In another embodiment, DNA pyrosequencing methylation assays are used for predicting HCC in body fluids such as plasma CF DNA by using CG IDs listed above, for example by using the below disclosed primers and standard conditions of pyrosequencing reactions:

25 cg02012576

Forward: GG TAGTTAGGAAGTTTAGAGGTTGTAGTA

Reverse (biotinylated): ACCACTACCCCAACCCAACCCTA

Sequence: GGTTTTAGGATGTTTG

cg03768777 (VASH2)

30 Forward: AGAATAATATTAGAGAATGGGATATGGAA

Reverse (biotinylated): ACAACTCCAAAATCCTACCT

Sequence: GAATGGGATATGGAATGA

cg05739190 (CCNJ)

Forward: GTTTAGGAGTTGGGTTTTAGTTGAG

Reverse (biotinylated): ACCCCACCCTAACTCCCTTACC

5 Sequence: TGGGTTTTAGTTGAGG

cg24804544 (GRID2IP)

Forward(biotinylated): GGGTAGGGGAGGGTTTTGAAATA

Reverse: TAACCCCCCTCCAACCTCATTC

Sequence: CACCCAACCTTCTCAAT

10 The specificity of the tissue of origin of the cancer is determined by measuring the DNA methylation of the following CGID cg02012576 (HPX)

Forward(biotinylated): ATTTTATGGGTATTAGTTTTAGGGAGAA

Reverse (biotinylated): CCAAACTATCCTATAACCTCTACAACTCA

Sequence: ACCATTACCACCCCT

15

In another embodiment, a polygenic multiplexed amplicon bisulfite sequencing DNA methylation assay is used for predicting cancer in body fluids such as plasma CF DNA by using CG IDs listed above. For example, predicting prostate cancer using the below disclosed primers and standard conditions that involve bisulfite conversion, sequential amplification with target specific primers (PCR 1) followed by barcoding primers (PCR 2) and multiplexed sequencing in a single next generation MiSeq sequencer (Illumina), demultiplexing using Illumina software, data extraction and quantification of methylation using standard methods for methylation analysis such as Methykit, followed by calculation of the weighted DNA methylation score and prediction of cancer from a small volume of body fluids such as plasma, saliva or urine.

The steps to detect prostate cancer the first PCR are performed as follows:

For CGID cg02879662

Forward primer:

30 5'ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGGTAGGAGTTTTGGG
AATTGG3'

Reverse primer:

5'gTgACTggAgTTCAgACgTgTgCTCTTCCgATCTCCACCCCTACAATCCCTAA3'

For CGID cg16232979

Forward primer:

5 5'ACACTCTTTCCCTACACgACgCTCTTCCgATCT
NNNNNYGGTTTTYGGGTTTTYGTATT3'

Reverse primer:

5'gTgACTggAgTTCAgACgTgTgCTCTTCCgATCTACRCAAAAATATAAATCRACR
ATC3'

10

To test that the cancer specifically originates in the prostate the first PCR is performed as follows:

For CGID: cg14041701 and cg14498227

Forward primer:

15 5'ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGTTTTGYGTTTTYGGA
TTTGGGTT3'

Reverse primer:

5'

gTgACTggAgTTCAgACgTgTgCTCTTCCgATCTCATAAACAAACACCTTTAAATAA

20

ACACTAAA3'

To barcode the samples, use a second PCR reaction with the following primers:

Forward primer:

5'AATgATACggCgACCACCgAgATCTACACTCTTTCCCTACACgAC3'

25

Barcoding primer (reverse):

5' CAAgCAgAAgACggCATACgAgATAGTCATCG gTgACTggAgTTCAgACgTg3'

(red bases are the index; 200 variations of this index are used)

In other embodiments, Receiver operating characteristics (ROC) assays are used
30 for detecting cancer by defining a threshold value between cancer and normal using

weighted DNA methylation measurements of CG IDs. Samples above/below threshold will be classified as cancer. For example, CGIDs listed above for detecting HCC:

cg02012576	cg03768777	cg05739190	cg24804544
------------	------------	------------	------------

5 In another embodiment, hierarchical Clustering analysis assays are used for predicting cancer by using measurements of methylation of CG IDs listed above.

 In another aspect of the inventive subject matter, methods for identifying DNA methylation markers for detecting cancer and other disease comprise the step of performing statistical analysis with the “binary-categorical differentiation (BCD)”
10 method previously disclosed regarding DNA methylation measurements obtained from clinical samples.

 In another embodiment, the method includes performing statistical analysis and the “binary-categorical differentiation (BCD)” method on DNA methylation measurements obtained from samples, with DNA methylation measurements obtained by
15 performing Illumina Beadchip 450K or EPIC array of DNA extracted from at least one sample.

 In another embodiment, the DNA methylation measurements are obtained by performing DNA pyrosequencing of DNA extracted from a sample followed by mass spectrometry based (Epityper™), PCR based methylation assays and targeted
20 amplification of a region spanning the target CG IDs disclosed herein from bisulfite converted DNA followed by barcoding in a second set of amplification and indexed-multiplexed sequencing on an Illumina next generation sequencer.

 In other embodiments, the statistical analysis includes Receiver operating characteristics (ROC) assays.

25 In other embodiments, the statistical analysis includes hierarchical clustering analysis assays.

Definitions

 As used herein, the term "CG" refers to a di-nucleotide sequence in DNA
30 containing cytosine and guanosine bases. These di-nucleotide sequences could become

5 methylated in human and other animals' DNA. The CG ID reveals its position in the human genome as defined by the Illumina 450K manifest (The annotation of the CGs listed herein is publicly available at <https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylation450k.db.html> and installed as an R package IlluminaHumanMethylation450k.db IlluminaHumanMethylation450k.db: Illumina Human Methylation 450k annotation data. R package version 2.0.9.).

10 As used herein, the term diagnostic equipment or equipment is any equipment known to those skilled in the art which can be used with one or more reagents to perform diagnostic tests such as those used for deriving DNA methylation measurements of the DNA methylation score according to embodiments of the claimed subject matter.

15 As used herein, the term "beta-value" refers to an estimation of methylation level at a CG ID position derived by normalization and quantification of Illumina 450K arrays using the ratio of intensities between methylated and unmethylated probes using the formula $\text{beta value} = \frac{\text{methylated C intensity}}{\text{methylated C intensity} + \text{unmethylated C intensity}}$ between 0 and 1 with 0 being fully unmethylated and 1 being fully methylated.

20 As used herein, the term "penalized regression" refers to a statistical method aimed at identifying the smallest number of predictors required to predict an outcome out of a larger list of biomarkers as implemented for example in the R statistical package "penalized" as described in Goeman, J. J., L1 penalized estimation in the Cox proportional hazards model. Biometrical Journal 52(1), 70-84.

As used herein, the term "clustering" refers to the grouping of a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

25 As used herein, the term "Hierarchical clustering" refers to a statistical method that builds a hierarchy of "clusters" based on how similar (close) or dissimilar (distant) are the clusters from each other as described for example in *Kaufman, L.; Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis (1 ed.). New York: John Wiley. ISBN 0-471-87876-6.*

30 As used herein, the term "Receiver operating characteristics (ROC) assay" refers to a statistical method that creates a graphical plot that illustrates the performance of a

predictor. The true positive rate of prediction is plotted against the false positive rate at various threshold settings for the predictor (i.e. different % of methylation) as described for example in Hanley, James A.; McNeil, Barbara J. (1982). "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve". Radiology 143 (1):
5 29–36.

As used herein, the term "multivariate or polygenic linear regression" refers to a statistical method that estimates the relationship between multiple "independent variables" or "predictors" such as percentage of methylation in CG IDs, and a "dependent variable" such as cancer. This method determines the "weight" or coefficient of each CG
10 IDs in predicting the "outcome" (dependent variable such as cancer) when several "independent variables" such as CG IDs are included in the model.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. shows a shortlist of completely unmethylated sites across blood samples and normal tissues in hundreds of individuals. Illustration A. shows CG IDs across 17 tissues that are unmethylated in all individuals (<0.1) in Illumina450K genome wide methylation arrays (GSE50192) were overlapped with unmethylated CG IDs in genome wide DNA methylation arrays of blood samples from 312 individuals (GSE61496) to generate a list of 33477 CG IDs. B illustrates shortlisting the most robustly unmethylated CG IDs the list of 33477 CG IDs from A was overlapped with unmethylated CG IDs in DNA methylation arrays of blood samples from 656 individuals (females and male) aged from 19 to 101 years (GSE40279). Combined, these analyses generated a list of high confidence 28754 CG IDs that are unmethylated across tissues and blood samples of many individuals across all age spans. These 28754 positions were used for discovery of sites that are categorically methylated in cancer but not in other tissues using the “binary-categorical differentiation (BCD)” method disclosed by the present inventive subject matter.

FIG. 2 is an illustration showing the Lack of Tissue specificity of current Circulating DNA markers for HCC. The illustrated heatmap shows 10 CG IDs shortlisted in Xu et al.,(Xu et al., 2017) as biomarkers of HCC and methylation levels for these sites in other normal tissues. Several of the CG IDs proposed as specific biomarkers of HCC are methylated in other tissues are well and show varying levels of methylation in blood DNA. (blue 0 methylation dark red 100% methylation)

FIG. 2A is an exploded view of a portion of FIG. 2 according to embodiments of the inventive subject matter.

FIG. 2B is an exploded view of a portion of FIG. 2 according to embodiments of the inventive subject matter.

FIG. 2C is an exploded view of a portion of FIG. 2 according to embodiments of the inventive subject matter.

FIG. 3 is an illustration showing the specificity of HCC DNA methylation markers discovered using the BCD method for cancer DNA. The illustrated heatmap shows 4 CG IDs selected as HCC DNA methylation markers by the BCD method described herein.

Methylation levels are categorically different between cancer (HCC) and normal tissues and blood, whereby the sites are unmethylated in all individuals in blood and other tissues and measurably methylated in HCC.

FIG. 3A is an exploded view of a portion of FIG. 3 according to embodiments of the
5 inventive subject matter.

FIG. 3B is an exploded view of a portion of FIG. 3 according to embodiments of the inventive subject matter.

FIG. 3C is an exploded view of a portion of FIG. 3 according to embodiments of the inventive subject matter.

10 FIG. 4 is an illustration showing the lack of cancer tissue of origin specificity of current DNA methylation markers for colorectal cancer and comparison with the “detect-spec” method according to embodiments of the inventive subject matter. Illustration A shows the CG sites in Sept9 gene as included in the “Epi-colon” CF DNA methylation marker for colorectal cancer (sold by Epigenomics Inc.) which can be used to detect many other
15 cancers utilizing methylation data from the TCGA collection of cancer DNA methylation data and thus lack specificity for colorectal cancer (HKG-Colon (HKG-epiCRC), blue). The markers disclosed in the present inventive subject matter for detection of colorectal cancer (Table 9) discovered using the BCD method (HKG-Colon orange) (Table 10) are highly specific for colorectal cancer when tested against other common solid tumors
20 cancers. Illustrations B and C are scatter-plots of DNA methylation values for tumor DNA from different individuals with different cancers using either the HKG-Colon (HKG-epiCRC) (B) or Epi-Colon (C) DNA methylation markers. Of note are the tight and categorical differences in DNA methylation between colorectal cancer and other cancers using the HKG-epiCRC markers (B) versus the scattered heterogenous profile of
25 Epi-Colon markers (C).

FIG. 4A is an exploded view of a portion of FIG. 4 according to embodiments of the inventive subject matter.

FIG. 4B is an exploded view of a portion of FIG. 4 according to embodiments of the inventive subject matter.

30 FIG. 5 is an illustration showing a discovery of a polygenic DNA methylation marker for early detection of liver cancer (HCC). Illustration A shows a table which lists the source

and number of patients whose methylation data was used for the discovery of a set of 4 CGIDs for detection of HCC according to an embodiment using the BCD method (Table 1) and CG IDs for determining the specific cancer of origin (Table 2). Illustration B at the bottom left panel of Fig. 5 (Detect) shows the combined methylation score for these CG IDs (Table 1) for each of the tested people listed from 1- 145 (79 normal and 66 HCC). The polygenic score categorically differentiates between people with HCC and normal liver tissue. Illustration C at the bottom right panel shows the methylation score for the 1 CGID detecting the specific tumor origin (Table 2) using data on 8 different tumors (Table 2). The markers categorically differentiate between cancers from other origins and HCC.

FIG. 6 is an illustration of a validation of a polygenic DNA methylation marker for HCC (spec) using DNA methylation data from GSE76269 (n=227). Illustration A is a ROC plot showing area under the curve for the HCC DNA methylation markers using 227 liver cancer patient DNA methylation data and 10 normal. Illustration B of Fig. 6 shows the sensitivity, specificity and accuracy of HCC detection. Illustration C shows the prediction rate of detection of HCC in the validation dataset.

FIG. 7 is an illustration of a validation of the polygenic HKG-epiLiver-detect and spec markers accuracy and specificity for HCC versus other cancers in TCGA methylation data (n=4166). Illustration A of FIG. 7 shows a detection rate of the HKG-Liver detect/spec markers DNA methylation data of patients with different cancers. Note the almost perfect specificity for HCC. B. ROC plot of the HKG-Liver-detect markers specificity and sensitivity for HCC in 4166 patient DNA methylation data from TCGA. C. Sensitivity and specificity to HCC versus cancers from other origins.

FIG. 8 is an illustration of a discovery of a polygenic DNA methylation marker for lung cancer. Illustration A of FIG. 8 is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of lung cancer disclosed in embodiments using the BCD method (Table 3) as well as CG IDs for determining the specific cancer tissue of origin (Table 4). Illustration B at the bottom left panel of FIG. 8 (Detect) shows the combined methylation score for these CG IDs (Table 3) for each of the tested people listed from 1- 20 (10 normal and 10 lung cancer). The polygenic score categorically differentiates between people with lung cancer and normal

tissue. Illustration C at the bottom right panel of FIG. 8 shows the methylation score for the CGIDs (Table 4) detecting the specific tumor origin using data from people who have 8 different tumors (n=80). In these embodiments, the markers categorically differentiate between cancers from other origins and lung cancer.

5 FIG. 9 is an illustration of the validation of the polygenic HKG-epiLung-detect and spec markers accuracy and specificity for HCC versus other cancers in TCGA methylation data (n=4166). Illustration A of FIG. 9 shows a detection rate of the HKG-epiLung detect/spec markers using DNA methylation data of patients with different cancers. Of note is the specificity for lung cancer. Illustration B of FIG. 9 shows a ROC plot of the
10 HKG-Lung-detect markers specificity and sensitivity for lung cancer on 4166 patient DNA methylation data from TCGA. Illustration C of FIG. 9 shows a sensitivity and specificity to lung cancer versus cancers from other origins.

FIG. 10 is an illustration of the discovery of a polygenic DNA methylation marker for prostate cancer. Illustration A of FIG. 10 is a table listing the source and number of a
15 patient whose methylation data was used for discovery of a set of CGIDs for detection of prostate cancer disclosed in embodiments using the BCD method (Table 5) and CGIDs for determining the specific cancer tissue of origin (Table 6). Illustration B at the bottom left panel of FIG. 10 (Detect) shows the combined methylation score for these CGIDs (Table 5) for each of the tested people listed from 1- 15 (5 normal and 10 prostate
20 cancer). The polygenic score categorically differentiates between people with prostate cancer and normal. Illustration C at the bottom right panel of FIG. 10 shows the methylation score for the CGs detecting the specific tumor tissue of origin (Table 6) using data from people who have 8 different tumors (n=80). In these embodiments, the markers categorically differentiate between cancers from other origins and prostate
25 cancer.

FIG. 11 is an illustration of a validation of the polygenic HKG-epiProstate-detect and spec markers accuracy and specificity for prostate versus other cancers in TCGA methylation data (n=4166). Illustration A of FIG. 11 shows the detection rate of the HKG-Prostate detect/spec markers using DNA methylation data from patients with
30 different cancers. Of note is the specificity for prostate cancer. Illustration B of FIG. 11 is a ROC plot of the HKG-Prostate-detect markers specificity and sensitivity for lung

cancer using DNA methylation data from 4166 patients in TCGA. Illustration C of FIG. 11 shows the sensitivity and specificity to prostate versus cancers from other origins. FIG. 12 is an illustration of a discovery of a polygenic DNA methylation marker for breast cancer. Illustration A of FIG. 12 is a table listing the source and number of patients
5 whose methylation data was used for discovery of a set of CGs for detection of breast cancer disclosed in embodiments using the BCD method (Table 7) and CGIDs for determining the specific cancer of origin (Table 8). Illustration B at the bottom left panel of FIG. 12 (Detect) shows the combined methylation score for these CGIDs (Table 7) for each of the tested people listed from 1- 27 (17 normal and 10 breast cancer). The
10 polygenic score categorically differentiates between people with breast cancer and normal breast tissue. Illustration C at the bottom right panel of FIG. 12 shows the methylation score for the CGIDs detecting the specific origin of the tumor (Table 8) using data from people who have 8 different tumors (n=80). In these embodiments, the markers categorically differentiate between cancers from other origins and breast cancer.
15 FIG. 13 is an illustration of a HKG-epiBreast-detect polygenic DNA methylation marker detect noninvasive as well as invasive breast cancer in validation cohort GSE60185 (n=285). Illustration A of FIG. 13 is a ROC plot showing area under the curve for the breast cancer polygenic DNA methylation marker using 239 breast cancer patient DNA methylation data, 17 mamoplastic surgery patients with no breast cancer and 29 adjacent
20 tissues. The sensitivity, specificity and accuracy for all breast cancer are listed in B and the prediction rate of DCIS (ductal carcinoma in situ), invasive and mixed breast cancer samples are shown in Illustration C of FIG. 13. Of note is that the breast cancer markers detect even very early breast cancers (DCIS).
FIG. 14 is an illustration of a validation of the polygenic HKG-epiBreast-detect and spec
25 markers accuracy and specificity for breast versus other cancers in TCGA methylation data (n=4166). Illustration A of FIG. 14 shows a detection rate of the HKG-epiBreast detect/spec markers in DNA methylation data from patients with different cancers. Of note is the specificity for breast cancer. Illustration B of FIG. 14 is a ROC plot of the HKG-Breast-detect markers specificity and sensitivity for detecting breast cancer using
30 DNA methylation data from 4166 patients in TCGA. Illustration C of FIG. 14 shows the sensitivity and specificity to breast versus cancers from other origins.

FIG. 15 is an illustration of a discovery of a polygenic DNA methylation marker for colorectal cancer (CRC). Illustration A of FIG. 15 is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of colorectal cancer disclosed in embodiments using the BCD method (Table 9) and CGIDs for determining the specific cancer of origin (Table 10). Illustration B at the bottom left panel of FIG. 15 (Detect) shows the combined methylation score for these CGIDs for each of the tested people listed from 1- 75 (25 normal and 50 colorectal cancer). The polygenic score categorically differentiates between people with cancer and normal tissue. Illustration C at the bottom right panel of FIG. 15 shows the methylation score for the CGIDs detecting the specific origin of the tumor using DNA methylation data from people who have 8 different tumors (n=80). In these embodiments, the markers categorically differentiate between cancers from other origins and colorectal cancer. FIG. 16 is an illustration validation of the polygenic HKG-epiCRC-detect and spec markers accuracy and specificity for CRC versus other cancers using TCGA DNA methylation data set (n=4166). Illustration A of FIG. 16 is a detection rate of the HKG-epiCRC detect/spec markers using DNA methylation data of patients with different cancers. Of note is the specificity for colorectal cancer. Illustration B of FIG. 16 is a ROC plot of the HKG-epiColon-detect markers specificity and sensitivity for colorectal cancer using DNA methylation data from 4166 patients in TCGA. Illustration C of FIG. 16 shows sensitivity and specificity to colorectal cancer versus cancers from other origins.

FIG. 17 illustrates a discovery of a polygenic DNA methylation marker for pancreatic cancer. Illustration A of FIG. 17 is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of pancreatic cancer disclosed in the present invention using the BCD method (Table 11) and CGIDs for determining the specific cancer of origin (Table 12). Illustration B at the bottom left panel of FIG. 17 (Detect) shows the combined methylation score for these CGIDs (Table 11) for each of the tested people listed from 1- 32 (12 normal and 20 pancreatic cancer). The polygenic score categorically differentiates between people with pancreatic cancer and normal tissue. Illustration C at the bottom right panel of FIG. 17 shows the methylation score for the CGIDs detecting the specific origin of the tumor

(Table 12) using data from people who have 10 different tumors (n=100). In these embodiments, the markers categorically differentiate between cancers from other origins and pancreatic cancer.

FIG. 18 is an illustration of a validation of the polygenic HKG-epiPancreas-detect and spec markers accuracy and specificity for pancreatic cancer versus other cancers in TCGA methylation data (n=4854). Illustration A of FIG. 18 is a detection rate of the HKG-epiPancreas detect/spec markers using DNA methylation data from patients with different cancers. Of note is the specificity for pancreatic cancer. Illustration B of FIG. 18 is a ROC plot of the HKG-epiPancreas-detect markers specificity and sensitivity for pancreatic cancer using DNA methylation data for 4854 patients in TCGA. Illustration C is the sensitivity and specificity to pancreatic cancer versus cancers from other origins. FIG. 19 is an illustration of a discovery of a polygenic DNA methylation marker for brain cancer (glioblastoma). Illustration A of FIG. 19 is a table listing the source and number of patient whose methylation data was used for discovery of a set of CGIDs for detection of brain cancer disclosed in the present invention using the BCD method (Table 13) and CGIDs for determining the specific origin of the cancer (Table 13). Illustration B at the bottom left panel (Detect/spec) shows the combined methylation score for these CGIDs (Table 13) for each of the tested people listed from 1- 16 (6 normal and 10 brain cancer). The polygenic score categorically differentiates between people with brain cancer, 110 other cancers, normal tissues.

FIG. 20 is an illustration of a validation of the polygenic HKG-epiBrain-detect and spec markers accuracy and specificity for breast versus other cancers in TCGA methylation data (n=4854). Illustration A is a detection rate of the HKG-epiBrain detect/spec markers using DNA methylation data from patients with different cancers. Note the specificity for brain cancer. Illustration B is a ROC plot of the specificity and sensitivity of the HKG-epiBrain-detect markers for brain cancer using DNA methylation data from 4854 patients in TCGA. Illustration C shows a sensitivity and specificity to brain cancer in TCGA data set (n=695).

FIG. 21 is an illustration of a discovery of a polygenic DNA methylation marker for detection of gastric (stomach) cancer. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for

detection of gastric cancer disclosed in the present invention using the BCD method (Table 14) and CGIDs for determining the specific origin of the cancer (Table 15).

Illustration B at the bottom left panel of FIG. 21 (Detect) shows the combined methylation score for these CGIDs (Table 14) for each of the tested people listed from 1- 28 (14 normal and 20 gastric cancer). The polygenic score categorically differentiates between people with gastric cancer and normal tissue. Illustration C at the bottom right panel of FIG. 21 (Spec) shows the polygenic methylation scores for people who have 10 different tumors (n=100). In these embodiments, the markers categorically differentiate between cancers from other origins and gastric cancer.

- FIG. 22 is an illustration of a validation of the polygenic HKG-Stomach-detect and spec markers accuracy and specificity for gastric cancer versus other cancers in TCGA methylation data (n=4817). Illustration A is a detection rate of the HKG-epiStomach detect/spec markers using DNA methylation data from patients with different cancers. Of note is the specificity for gastric cancer. Illustration B is a ROC plot specificity and sensitivity of the HKG-epiStomach-detect spec 1 markers for stomach (gastric cancer) using DNA methylation data from 4420 patients in TCGA. Illustration C is a ROC plot of specificity and sensitivity the HKG-epiStomach- spec 1 markers for gastric cancer using DNA methylation data from 4854 patients in TCGA. Of note is that there is a significant cross reactivity with colorectal and esophageal cancer which attests to a shared origin.
- FIG. 23 is an illustration of a discovery of a polygenic DNA methylation marker for ovarian cancer. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of ovarian cancer disclosed in the present invention using the BCD method (Table 16) and CGIDs for determining the specific cancer of origin (Table 17). Illustration B at the bottom left panel of FIG. 23 (Detect) shows the combined methylation score for these CGIDs for each of the tested people listed from 1- 15 (5 normal and 10 ovarian cancer). The polygenic score categorically differentiates between people with ovarian cancer and normal tissue. Illustration C at the bottom right panel of FIG. 23 shows the methylation score for the CGIDs detecting the specific tumor origin using data from people who have 11 different tumors (n=110). In these embodiments, the markers categorically differentiate between cancers from other origins and ovarian cancer.

FIG. 24 is an illustration of a validation of the polygenic HKG-epiOvarian-detect and spec markers accuracy and specificity for cervical versus other cancers in TCGA methylation data (n=6522). Illustration A is a detection rate of the HKG-epiOvarian detect/spec markers using DNA methylation data from patients with different cancers.

- 5 Note the specificity for ovarian cancer. Illustration B is a ROC plot specificity and sensitivity of the HKG-epiOvarian-detect and spec markers for ovarian cancer on DNA methylation data from 4723 patients in TCGA. Illustration C shows the sensitivity and specificity to ovarian cancer.

- FIG. 25 is an illustration of a discovery of a polygenic DNA methylation marker for cervical cancer. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of cervical cancer disclosed in the present invention using the BCD method (Table 18) and CGs for determining the specific cancer of origin (Table 19). Illustration B at the bottom left panel of FIG. 25 (Detect) shows the combined methylation score for these CGIDs (Table 18) for each of the tested people listed from 1- 30 (20 normal and 10 ovarian cancer). The polygenic score categorically differentiates between cervical cancer and normal tissue. Illustration C at the bottom right panel of FIG. 25 shows the methylation score for the CG IDs detecting the specific origin of the tumor (Table 19) using data from people who have 8 different tumors (n=80). In these embodiments, the markers categorically differentiate between cancers from other origins and cervical cancer, note however some measurable detection of colorectal cancer.

- FIG. 26 is an illustration of a validation of the polygenic HKG-Cervix-detect and spec markers accuracy and specificity for cervix versus other cancers in TCGA methylation data (n=6522). Illustration A shows a detection rate of the HKG-Cervix detect/spec markers using DNA methylation data from patients with different cancers. Of note is the specificity for cervical cancer. Illustration B is a ROC plot of specificity and sensitivity of the HKG-Cervix-detect spec markers for cervical cancer using DNA methylation data from 4420 patients in TCGA. Illustration C shows a sensitivity and specificity to cervical cancer.

- 30 FIG. 27 is an illustration of a discovery of a polygenic DNA methylation marker for Head and Neck squamous cell Carcinoma (HNSC). Illustration A is a table listing the source

and number of patients whose methylation data was used for discovery of a set of CG IDs for detection of HNSC disclosed in the present invention using the BCD method (Table 20) and CGs for determining the specific cancer of origin (Table 21). Illustration B at the bottom left panel of FIG. 27 shows the combined methylation score for these CG IDs (Table 20) for each of the tested people listed from 1-140 (10 cancer, 10 normal and 120 other cancers). Illustration C shows the polygenic score which categorically differentiates between HNSC and normal tissue samples in the embodiments as well as categorically differentiating between cancers from other origins and HNSC.

FIG. 28 is an illustration of a validation of the polygenic HKG-epiHNSC -detect/spec markers accuracy and specificity for HNSC versus other cancers in TCGA methylation data (n=4166). Illustration A is a detection rate of the HKG-epiHNSC detect/spec markers using DNA methylation data from patients with different cancers. Of note is the specificity for HNSC. Illustration B is a ROC plot of the specificity and sensitivity of HKG-epiHNSC-detect markers for HNSC on DNA methylation data from 4166 patients in TCGA. Illustration C shows the sensitivity and specificity for HNSC versus cancers from other origins.

FIG. 29 is an illustration of a discovery of a polygenic DNA methylation marker for esophageal cancer. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of esophageal cancer disclosed in embodiments using the BCD method (Table 22) and CGIDs for determining the specific origin of the cancer (Table 23). Illustration B at the bottom left panel of FIG. 29 shows the combined methylation score for these CGIDs (Table 22) for each of the tested people listed from 1-15 (6 normal, 10 cancer). Illustration C shows the polygenic score categorically differentiating between esophageal cancer and normal tissue in the embodiments as well as categorically differentiating between cancers from other origins and esophageal cancer listed from 1-220 (20 cancer, 190 other cancer and 10 healthy blood).

FIG. 30 is an illustration of a validation of the polygenic HKG-epiEsophageal -detect/spec markers accuracy and specificity for esophageal cancer versus other cancers in TCGA methylation data (n=7102). Illustration A shows a detection rate of the HKG-epiEsophageal detect/spec markers using DNA methylation data from patients with

different cancers. Of note is the specificity for esophageal cancer. Illustration B is a ROC plot of the specificity and sensitivity of the HKG-epiEsophageal-detect markers for HNSC on 4166 patient DNA methylation data in TCGA. Illustration C shows sensitivity and specificity to esophageal cancer versus cancers from other origins.

- 5 FIG. 31 is an illustration of a discovery of a polygenic DNA methylation marker for bladder cancer. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of bladder cancer disclosed in embodiments using the BCD method (Table 24) and CGIDs for determining the specific origin of the cancer (Table 25). Illustration B at the bottom left panel of FIG.
- 10 31 (Detect) shows the combined methylation score for these CGIDs (Table 24) for each of the tested people listed from 1- 15 (5 normal and 10 bladder cancer). Illustration C at the bottom right panel of FIG. 31 shows the methylation score for the CGIDs (Table 25) detecting the specific origin of the tumor using data from people who have 13 different tumors (n=130). In these embodiments, the markers differentiate between cancers from
- 15 other origins and bladder cancer. Also of note are some measurable detection of colorectal cancer with these markers.

- FIG. 32 is an illustration of a validation of accuracy and specificity of the polygenic HKG-epiBladder-detect and spec markers for bladder cancer versus other cancers in TCGA (n=4723). Illustration A shows a detection rate of the HKG-epiBladder spec (A)
- 20 and detect markers (B) on DNA methylation data of patients with different cancers (A) and bladder cancer (B). Illustration C is a ROC plot of the specificity and sensitivity of HKG-epiBladder spec markers for bladder cancer using DNA methylation data from 4420 patients in TCGA. Illustration D is a ROC plot of the specificity and sensitivity of the HKG-epiBladder detect markers for bladder cancer (n=440).

- 25 FIG. 33 is an illustration of a discovery of a polygenic DNA methylation marker for kidney cancer. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of kidney cancer disclosed in embodiments using the BCD (hypo) method and for determining the specific origin of the cancer (Table 26). Illustration B at the bottom left panel of FIG. 33
- 30 (Detect/spec) shows the combined methylation score for these CGIDs (Table 26) for each of the tested people listed from 1- 226 (180 other cancers, 10 healthy blood, 6 normal

kidney, 30 renal cancer). In these embodiments, the polygenic score categorically differentiates between kidney cancer, other cancers and normal blood.

FIG. 34 is an illustration of a validation of the accuracy and specificity of polygenic HKG-epiKidney-detect and spec markers for kidney cancer versus other cancers and normal tissues using TCGA DNA methylation data (n=7102). Illustration A is the detection rate of the HKG-epiKidney detect/spec markers using DNA methylation data from different cancers. Of note is the specificity for kidney cancer. Illustration B is a ROC plot of the specificity and sensitivity of HKG-Cervix-detect spec markers for renal cancer using DNA methylation data from 6367 cancers in TCGA. Illustration C is a sensitivity and specificity to renal (kidney) cancer. Of additional note is a crossover with brain, HCC and testicular cancer.

FIG. 35 is an illustration of a discovery of a polygenic DNA methylation marker for testicular cancer. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of testicular cancer disclosed in embodiments using the BCD (hypo) method and for determining the specific cancer of origin (Table 27). Illustration B at the bottom left panel of FIG. 35 (Detect/spec) shows the combined methylation score for these CG IDs (Table 27) for each of the tested people listed from 1- 226 (10 testicular cancer, 180 other cancers, 10 normal blood). In these embodiments, the polygenic score categorically differentiates between testicular cancer and normal blood and other cancers.

FIG. 36 is an illustration of a validation of the accuracy and specificity of the polygenic HKG-epiTestis-detect and spec markers for testicular cancer versus other normal tissues and cancers in TCGA methylation data (n=7102). Illustration A shows the detection rate of the HKG-epiTestis detect/spec markers using DNA methylation data from patients with different cancers. Of note is the specificity for Testis cancer. Illustration B is a ROC plot of the specificity and sensitivity of HKG-epiTestis-detect spec markers for testicular cancer using DNA methylation data from 6367 patients in TCGA. Illustration C is the sensitivity and specificity to testicular cancer.

FIG. 37 is an illustration of a discovery of a Pan cancer polygenic DNA methylation marker for 13 common cancers. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection

of 13 common cancers (Table 28) (bladder cancer, brain cancer, breast cancer, cervical cancer, colorectal cancer CRC, esophageal cancer, liver cancer, lung cancer, ovarian cancer, pancreatic cancer, prostate cancer, and stomach cancer) disclosed in embodiments using the BCD method (Table 28). Illustration B shows the combined methylation score
 5 for these CGIDs for each of the tested people listed from 1-310 (180 cancer and 10 normal). In these embodiments, the polygenic score differentiates between cancers and normal tissue.

FIG. 38 is an illustration of a validation of the polygenic HKG epiPancancer markers accuracy and specificity in TCGA methylation data (n=7102). Illustration A shows
 10 Methylation scores calculated using the epiPancancer polygenic DNA methylation markers in patients with 13 different cancers using TCGA data. Illustration B is a ROC plot of the specificity and sensitivity of HKG-epiPancancer detect and spec markers using DNA methylation data from for all cancers from 4878 patients in TCGA. Illustration C is a ROC plot of the epiPancancer polygenic markers describing specificity and sensitivity
 15 for detection of 13 common cancers. Illustration D shows the overall sensitivity and specificity of the pan cancer markers for detecting cancer. In these embodiments, one or more colors are used, for example orange (weighted methylation score) and blue (detection of one BCD marker per sample is scored as a positive cancer).

FIG 39 is an illustration of a discovery of polygenic DNA methylation marker for
 20 Melanoma. Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of melanoma (Table 45) disclosed in embodiments using the BCD method (Table 45). Illustration B is the combined methylation score for these CGIDs for each of the tested people listed from 1-220 (other cancers and normal blood) and 10 patients with melanoma. In these
 25 embodiments, the polygenic score differentiates between melanoma, other cancers and normal tissue.

FIG. 40 is an illustration of a validation of the accuracy and specificity of the polygenic HKG-epiMelanoma-detect and spec markers for melanoma versus other normal tissues and cancers in TCGA methylation data (n=7102). Illustration A shows the detection rate
 30 of the HKG-epiMelanoma detect/spec markers using DNA methylation data from patients with different cancers. Of note is the specificity for melanoma (with overlapping

detection of liver cancer brain and cancer and prostate cancer). Illustration B is a ROC plot of the specificity and sensitivity of HKG-Melanoma-detect spec markers for melanoma using DNA methylation data from 6367 patients in TCGA. Illustration C shows the sensitivity and specificity to melanoma.

- 5 FIG 41 is an illustration of a discovery of polygenic DNA methylation marker for blood cancers (Acute Myeloid Leukemia (AML). Illustration A is a table listing the source and number of patients whose methylation data was used for discovery of a set of CGIDs for detection of blood cancer AML (Table 46) disclosed embodiments using the BCD method (Table 46). Illustration B is the combined methylation score for these CGIDs for
- 10 each of the tested people listed from 1-10 (normal blood) and 10 patients with AML. In these embodiments, the polygenic score differentiates between AML and normal blood. FIG. 42 is an illustration of a validation of the accuracy and specificity of the polygenic HKG-epiAML-detect and spec markers for AML in GSE86409 (n=79) and in TCGA (n-140) versus normal blood in GSE40279 and GSE61496 (n=968). Illustration A shows a
- 15 detection rate of the HKG-epiAML detect/spec markers using DNA methylation data from patients with AML and healthy blood. Of note is the specificity for melanoma (with overlapping detection of liver cancer brain and cancer and prostate cancer). Illustration B is a ROC plot of the specificity and sensitivity of HKG-AML-detect spec markers for AML using DNA methylation data from GSE86409 (n=79), TCGA (n-140) GSE40279
- 20 and GSE61496 (n=968). Illustration C shows the sensitivity and specificity to AML. FIG. 43 is an illustration of a validation that the primers selected for detecting different cancers exhibit BCD properties ~0 methylation in plasma derived from normal people (each sample is a mixture of plasma from normal patients). The first PCR1 reaction targeting the specific CGs was performed using sequence targeted primers. Following a
- 25 second PCR, the amplified fragments were purified and subjected to next generation sequencing. DNA methylation was quantified in each of the indicated CG ID positions. FIG. 44 is an illustration of a validation that the indicated primers selected for detecting different cancers exhibit BCD properties ~0 methylation in plasma derived from normal people (each sample is a mixture of plasma from normal patients).
- 30 FIG. 45 is an illustration of a primer design for multiplex amplification and sequencing. The first PCR reaction targets the specific regions of interest, note PCR1 primer has

complementary sequence to second PCR2 primers. The second set of primers introduces the index for each patient as well as the reverse and forward sequencing primers.

FIG. 46 is an illustration of an optimization of PCR conditions for detecting prostate cancer. A multiplex PCR1 reaction using varying primer concentrations as indicated

- 5 DNA for the three markers of prostate cancer HIF3A 232 bp, TPM4 213 bp, and CTTN 199 bp is shown on the right panel.

FIG. 47. is an illustration of a bioinformatics workflow for determining DNA methylation levels. PCR2 products are combined, quantified and purified and subjected to next generation sequencing on a Miseq Illumina sequencer. Sequence is demultiplexed,

- 10 FASTQ files are generated for each patient and analyzed with the workflow shown in the scheme. DNA methylation scores are calculated for each patient.

DETAILED DESCRIPTION

All illustrations of the drawings are for the purpose of describing selected embodiments and are not intended to limit the scope of the claimed subject matter.

5

Embodiment 1. "Discovery of categorically unmethylated CGIDs across hundreds of individuals in normal tissues and blood DNA

Cell free DNA originating in tumors is known to be found in body fluids such as plasma, urine and in feces. It is also established that DNA methylation profiles of CF
10 tumor DNA are similar to tumor DNA (Dominguez-Vigil et al., 2018). A vast body of data has established that tumor DNA is differentially methylated compared to normal tissues (Luczak & Jagodzinski, 2006). Therefore, many groups have tried to delineate by logistic regressions CGID positions in DNA (CG IDs in the Illumina 450K manifest) that are differentially methylated between cancerous and its normal tissue of origin for
15 example, liver cancer versus adjacent liver tissue. However, since these methods measure quantitative differences between cancer and untransformed tissue rather than categorical qualitative differences, these quantitative differences between tumor and normal tissue would be diluted and erased by CF DNA from normal tissue, leading to false negatives and reduced sensitivity. In addition, other tissues that were not included in the analysis
20 might have a DNA methylation profile similar to tumor DNA and since most studies only compare the tumor DNA to its untransformed counterparts and not to other tissues, this could lead to false positives. Varying and unpredictable quantities of DNA from different tissues have been detected in CF DNA (Breitbach et al., 2014) and thus the measured DNA methylation reflects a composite of unknown and unpredictable mixture of tissue
25 DNA from different sources and tumor DNA. Thousands of tumor samples have been subjected to genome wide DNA methylation analysis using Illumina 450K arrays and are found in the public domain (TCGA). Examining the profiles of methylation of many normal tissues as well as cancer tissues, the inventors noticed that there is a significant group of CGs in the genome that are completely unmethylated in all normal tissues but
30 methylated in DNA from tumors. A subset of these sites is unmethylated across numerous individuals whose DNA methylation was profiled in the public domain. The

inventors also noticed that in many cancers these robustly unmethylated sites become methylated in cancer. Thus, creating a qualitative “categorical difference” between tumor DNA and all other DNA that might be found in blood. Using deep next generation sequencing even few methylated molecules could be easily identified on a background of
 5 completely unmethylated copies.

Data bases; Illumina 450K DNA methylation data

We used publicly available data bases of normalized beta values of methylation for ~450,000 CG across the human genome from a large number of individuals deposited
 10 either in the Gene expression Omnibus (GEO) <https://www.ncbi.nlm.nih.gov/geo/> or The Cancer Genome Atlas TCGA <https://cancergenome.nih.gov/> public data bases. We used the following databases to derive the list of robustly unmethylated CG IDs in many normal tissues and blood DNA: GSE50192, GSE50192, GSE40279.

DNA from white blood cells is one of the main sources of CF DNA in plasma.

15 The inventors first generated a list of 47981 CGIDS that are unmethylated in all individuals in 17 different somatic human tissues using Illumina 450K data in GSE50192 and the logical COUNTIF and IF functions in Excel:

$NmCGID_x = COUNTIF(betaCGID_{x:n_1:n_i}, ">0.1")$

$umCGID_x = IF(NmCGID_x=0, TRUE, FALSE)$

20 $NmCGID_x$ =number of normal subjects that have the $CGID_x$ methylated.

$umCGID_x$ =unmethylated $CGID_x$ in all subjects

$betaCGID_x$ = the methylation values for a given $CGID_x$

x = any CGID on the Illumina 450k array

n_1 =the first subject in the array,

25 n_i =the last subject in the array.

The inventors then generated a list of 68260 unmethylated CGIDs (UMCGIDs) in blood DNA from 312 individuals using the same criteria. The inventors then overlapped the list of 47981 and 68260 CG IDS and obtained a list of 33477 CG IDs that are
 30 unmethylated in both blood and somatic tissues across all individuals (Fig. 1A). To increase the robustness of this list of unmethylated CG IDs the inventors delineated a list

of 60,379 CG IDs unmethylated CGIDs in Illumina 450K arrays of whole blood DNA from 656 individuals males and females aged from 19 to 101 years (GSE40279). These are robustly unmethylated sites in blood that are sex and age independent across hundreds of individuals. This list of 60,379 CG IDs was overlapped with the list of 33,477 CG IDs that are unmethylated both in somatic tissues and blood to generate a final list of 28,754 CG IDs which were used for discovery of categorical methylation markers for cancer. This list includes CG ID positions that are robustly unmethylated across tissues and individuals.

To identify DNA methylation positions that are categorically different between cancer and normal tissues the inventors examined whether any of these 28754 CG IDs are methylated in different cancers. The inventors noticed following examination of tumor DNA methylation data that methylation of a subset of these 28754 CG IDs is common in tumor DNA from individual patients. However, not all individuals have the same position methylated. Thus, a combination of CG IDs is required to detect cancer with high specificity. The inventors therefore discovered a polygenic combination of CG IDs for detection of cancers.

The inventors used 10 to 50 DNA methylation profiles from the public domain from either TCGA or GEO as a “discovery set” to discover a polygenic set of CGIDs whose methylation state is “categorically” different between tumor and normal tissues that could detect cancer with highest sensitivity and specificity. These CGIDs were then tested on hundreds of TCGA and GEO tumor DNA methylation array data as a “validation set” to validate the sensitivity and specificity of the polygenic DNA methylation markers for detecting cancer as disclosed in Embodiment 2.

Embodiment 2: Binary-categorical differentiation (BCD)” method for detecting cancer in cell free DNA.

The following publicly available data bases of normalized beta values of methylation for ~450,000 CGs (CG IDs) across the human genome were used to derive the list of cancer specific DNA methylation markers:

Table 29 liver cancer

Disease Status	Source	Cohort	N	Detect/Spec
Normal Liver	GSE61258	Discovery	79	Detect
HCC	TCGA	Discovery	50	Detect
HCC	TCGA	Discovery	10	Spec
Non-HCC cancers	TCGA	Discovery	80	Spec
HCC	GSE75041	Validation	66	Detect
HCC	GSE76269	Validation	227	Detect
HCC	TCGA	Validation	430	Detect
Bladder cancer	TCGA	Validation	439	Spec
Brain cancer	TCGA	Validation	689	Spec
Breast cancer	TCGA	Validation	891	Spec
Cervix cancer	TCGA	Validation	312	Spec
CRC cancer	TCGA	Validation	459	Spec
ESCA cancer	TCGA	Validation	202	Spec
Kidney cancer	TCGA	Validation	871	Spec
Lung cancer	TCGA	Validation	919	Spec
Ovarian cancer	TCGA	Validation	10	Spec
PANC cancer	TCGA	Validation	195	Spec
PRAD cancer	TCGA	Validation	553	Spec
Stomach cancer	TCGA	Validation	397	Spec
Testis cancer	TCGA	Validation	156	Spec
Normal Liver	GSE76269	Validation	10	Spec
Normal Liver	GSE69852	Validation	6	Spec
Normal Liver	GSE75041	Validation	10	Spec
Normal Blood	GSE40279	Validation	656	Spec
Normal Blood	GSE61496	Validation	312	Spec

Table 30 lung cancer

Disease Status	Source	Cohort	N	Detect/Spec
Lung Cancer	GSE63704	Discovery	10	Detect
Normal Lung	GSE66836	Discovery	10	Detect
Lung Cancer	TCGA	Discovery	10	Spec
Non-lung cancers	TCGA	Discovery	80	Spec
Lung Cancer	GSE66836	Validation	164	Detect
Lung Cancer	GSE63704	Validation	17	Detect
Lung Cancer	GSE76269	Validation	56	Detect
Lung Cancer	TCGA	Validation	919	Spec
Bladder Cancer	TCGA	Validation	439	Detect
Brain Cancer	TCGA	Validation	689	Spec
Breast Cancer	TCGA	Validation	891	Spec
Cervical Cancer	TCGA	Validation	312	Spec
CRC	TCGA	Validation	459	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney Cancer	TCGA	Validation	871	Spec
Ovarian Cancer	TCGA	Validation	10	Spec
PANC	TCGA	Validation	195	Spec
PRAD	TCGA	Validation	553	Spec
Stomach Cancer	TCGA	Validation	396	Spec
Testis Cancer	TCGA	Validation	156	Spec
Normal Lung	GSE63704	Validation	112	Detect

Table 31 prostate cancer

Reference/Status	Source	Category	N	Detect/Spec
Control	GSE52955	Discovery	5	Detect
PRAD	TCGA	Discovery	10	Detect
PRAD	TCGA	Discovery	10	Spec
Non-PRAD	TCGA	Discovery	80	Spec
PRAD	GSE73549	Validation	77	Detect
PRAD	GSE52955	Validation	25	Detect
PRAD	TCGA	Validation	553	Spec
Bladder	TCGA	Validation	439	Spec
Brain	TCGA	Validation	689	Spec
Breast	TCGA	Validation	891	Spec
Cervix	TCGA	Validation	312	Spec
CRC	TCGA	Validation	459	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney	TCGA	Validation	871	Spec
Lung	TCGA	Validation	919	Spec
Ovarian	TCGA	Validation	10	Spec
PANC	TCGA	Validation	195	Spec
Stomach	TCGA	Validation	397	Spec
Testis	TCGA	Validation	156	Spec
Normal Prostate	GSE73549	Validation	15	Detect
Normal Prostate	GSE52955	Validation	5	Detect
Normal Blood	GSE40279	Validation	656	Spec
Normal Blood	GSE61496	Validation	312	Spec

Table 32 breast cancer

Reference/Status	Source	Category	N	Detect/Spec
Healthy	GSE60185	Discovery	17	Detect
Breast Cancer	TCGA	Discovery	10	Detect
Breast Cancer	TCGA	Discovery	10	Spec
Non-breast cancers	TCGA	Discovery	80	Spec
Breast Cancer	GSE60185	Validation	239	Detect
Breast Cancer	GSE75067	Validation	188	Detect
Breast Cancer	TCGA	Validation	891	Detect
Bladder Cancer	TCGA	Validation	439	Spec
Brain Cancer	TCGA	Validation	689	Spec
Cervical Cancer	TCGA	Validation	312	Spec
CRC	TCGA	Validation	459	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney Cancer	TCGA	Validation	871	Spec
Lung Cancer	TCGA	Validation	919	Spec
Ovarian Cancer	TCGA	Validation	10	Spec
PANC	TCGA	Validation	195	Spec
PRAD	TCGA	Validation	553	Spec
Stomach Cancer	TCGA	Validation	396	Spec
Testis Cancer	TCGA	Validation	156	Spec
Healthy breast tissue	GSE101961	Validation	121	Detect
Healthy breast tissue	GSE60185	Validation	17	Detect
Normal adjacent tiss	GSE60185	Validation	29	Detect
Normal Blood	GSE40279	Validation	656	Spec
Normal Blood	GSE61496	Validation	312	Spec

Table 33 colorectal cancer CRC				
Disease Status	Source	Cohort	N	Detect/Spec
Control	GSE6550	Discovery	25	Detect
CRC	TCGA	Discovery	50	Detect
CRC	TCGA	Discovery	10	Spec
Non-HCC	TCGA	Discovery	80	Spec
CRC	TCGA	Validation	260	Detect
HCC	TCGA	Validation	459	Detect
Bladder	TCGA	Validation	439	Spec
Brain	TCGA	Validation	689	Spec
Breast	TCGA	Validation	891	Spec
Cervix	TCGA	Validation	312	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney	TCGA	Validation	871	Spec
Lung	TCGA	Validation	919	Spec
Ovarian	TCGA	Validation	10	Spec
PANC	TCGA	Validation	195	Spec
PRAD	TCGA	Validation	553	Spec
Stomach	TCGA	Validation	397	Spec
Testis	TCGA	Validation	156	Spec
Normal Colorectal	GSE6550	Validation	8	Detect
Normal Blood	GSE40279	Validation	656	Spec
Normal Blood	GSE61496	Validation	312	Spec

Table 34 Pancreatic cancer				
Disease Status	Source	Cohort	N	Detect/Spec
Healthy	GSE53051	Discovery	12	Detect
Pancreatic Cancer	TCGA	Discovery	20	Detect
Pancreatic Cancer	TCGA	Discovery	20	Spec
Non-pancreatic	TCGA	Discovery	100	Spec
Pancreatic Cancer	TCGA	Validation	24	Detect
Pancreatic Cancer	TCGA	Validation	195	Detect
Bladder Cancer	TCGA	Validation	439	Spec
brain Cancer	TCGA	Validation	689	Spec
Breast Cancer	TCGA	Validation	891	Spec
Cervical Cancer	TCGA	Validation	312	Spec
CRC	TCGA	Validation	459	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney Cancer	TCGA	Validation	871	Spec
Lung Cancer	TCGA	Validation	919	Spec
Ovarian Cancer	TCGA	Validation	10	Spec
PRAD	TCGA	Validation	553	Spec
Stomach Cancer	TCGA	Validation	396	Spec
Testis Cancer	TCGA	Validation	156	Spec
Normal Blood	GSE40179	Validation	656	Spec
Normal Blood	GSE61496	Validation	312	Spec

Table 35 Brain cancer

Disease Status	Source	Cohort	N	Probes / Spec.
Brain Cancer	TCGA	Discovery	10	Spec-Detect
Non-brain cancer	TCGA	Discovery	158	Spec-Detect
Brain Cancer	GSE36278	Validation	136	Spec-Detect
Brain Cancer	GSE58298	Validation	40	Spec-Detect
Brain Cancer	GSE58218	Validation	228	Spec-Detect
Brain Cancer	TCGA	Validation	689	Spec-Detect
Bladder Cancer	TCGA	Validation	439	Spec-Detect
Breast Cancer	TCGA	Validation	891	Spec-Detect
Cervical Cancer	TCGA	Validation	312	Spec-Detect
CRC	TCGA	Validation	459	Spec-Detect
ESCA	TCGA	Validation	202	Spec-Detect
HCC	TCGA	Validation	430	Spec-Detect
Kidney Cancer	TCGA	Validation	871	Spec-Detect
Lung Cancer	TCGA	Validation	919	Spec-Detect
Ovarian Cancer	TCGA	Validation	10	Spec-Detect
PANC	TCGA	Validation	195	Spec-Detect
PRAD	TCGA	Validation	553	Spec-Detect
Stomach Cancer	TCGA	Validation	396	Spec-Detect
Testis Cancer	TCGA	Validation	156	Spec-Detect
Brain Cancer	GSE36278	Validation	6	Spec-Detect
Normal Blood	GSE40279	Validation	656	Spec-Detect
Normal Blood	GSE61496	Validation	312	Spec-Detect

Table 36 Stomach cancer

Disease Status	Source	Cohort	N	Probes / Spec.
Healthy	GSE99553	Discovery	14	Detect
Stomach Cancer	TCGA	Discovery	20	Detect
Stomach Cancer	TCGA	Discovery	7	Spec
Non-stomach cancer	TCGA	Discovery	100	Spec
Stomach Cancer	TCGA	Validation	396	Detect
Bladder Cancer	TCGA	Validation	439	Spec
Brain Cancer	TCGA	Validation	689	Spec
Breast Cancer	TCGA	Validation	891	Spec
Cervical Cancer	TCGA	Validation	312	Spec
CRC	TCGA	Validation	459	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney Cancer	TCGA	Validation	871	Spec
Lung Cancer	TCGA	Validation	919	Spec
Ovarian Cancer	TCGA	Validation	10	Spec
PANC	TCGA	Validation	195	Spec
PRAD	TCGA	Validation	553	Spec
Testis Cancer	TCGA	Validation	156	Spec
Normal Tissue	GSE99553	Validation	42	Detect
Normal Blood	GSE40179	Validation	656	Spec
Normal Blood	GSE61496	Validation	312	Spec

Table 37 Ovarian cancer			
Disease Status	Source	Colony	n
Healthy	GSE65821	Discovery	6 Detect
Ovarian Cancer	TCGA	Discovery	10 Detect
Ovarian Cancer	TCGA	Discovery	10 Spec
Non-ovarian cancers	TCGA	Discovery	110 Spec
Ovarian Cancer	GSE65821	Validation	113 Detect
Bladder Cancer	TCGA	Validation	439 Spec
Brain Cancer	TCGA	Validation	689 Spec
Breast Cancer	TCGA	Validation	891 Detect
Cervical Cancer	TCGA	Validation	312 Spec
CRC	TCGA	Validation	459 Spec
ESCA	TCGA	Validation	202 Spec
HCC	TCGA	Validation	430 Spec
Kidney Cancer	TCGA	Validation	871 Spec
Lung Cancer	TCGA	Validation	919 Spec
PANC	TCGA	Validation	195 Spec
PRAD	TCGA	Validation	553 Spec
Stomach Cancer	TCGA	Validation	396 Spec
Testis Cancer	TCGA	Validation	156 Spec
Healthy ovarian tissue	GSE87621	Validation	9 Detect
Healthy ovarian tissue	GSE74845	Validation	216 Detect
Healthy ovarian tissue	GSE81228	Validation	10 Detect
Normal Blood	GSE40279	Validation	656 Spec

Table 38 Cervical cancer			
Disease Status	Source	Colony	n
Healthy	GSE46306	Discovery	20 Detect
Cervical Cancer	TCGA	Discovery	10 Detect
Cervical Cancer	TCGA	Discovery	10 Spec
Non-cervical cancers	TCGA	Discovery	110 Spec
Cervical Cancer	GSE68339	Validation	270 Detect
Cervical Cancer	TCGA	Validation	312 Detect
Bladder Cancer	TCGA	Validation	439 Spec
Brain Cancer	TCGA	Validation	689 Spec
Breast Cancer	TCGA	Validation	891 Spec
CRC	TCGA	Validation	459 Spec
ESCA	TCGA	Validation	202 Spec
HCC	TCGA	Validation	430 Spec
Kidney Cancer	TCGA	Validation	871 Spec
Lung Cancer	TCGA	Validation	919 Spec
Ovarian Cancer	TCGA	Validation	10 Spec
PANC	TCGA	Validation	195 Spec
PRAD	TCGA	Validation	553 Spec
Stomach Cancer	TCGA	Validation	396 Spec
Testis Cancer	TCGA	Validation	156 Spec
Normal Blood	GSE40279	Validation	656 Spec
Normal Blood	GSE61496	Validation	312 Spec

Table 39 HNSC

Disease Status	Source	Cohort	N	Detect/Spec
HNSC	TCGA	Discovery	20	Detect
Control	GSE75537	Discovery	10	Detect
HNSC	TCGA	Discovery	20	Spec
Non-HNSC	TCGA	Discovery	190	Spec
Healthy	GSE40279	Discovery	10	Spec
HNSC	TCGA	Validation	580	Detect
Control	GSE52068	Validation	24	Detect
HNSC	GSE75537	Validation	54	Detect
HNSC	GSE79556	Validation	83	Detect
Control	GSE75537	Validation	54	Detect
HNSC	GSE52068	Validation	24	Detect
Bladder	TCGA	Validation	439	Spec
Brain	TCGA	Validation	689	Spec
Breast	TCGA	Validation	891	Spec
Cervix	TCGA	Validation	312	Spec
HCC	TCGA	Validation	459	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney	TCGA	Validation	871	Spec
Lung	TCGA	Validation	919	Spec
Ovarian	TCGA	Validation	10	Spec
PANC	TCGA	Validation	195	Spec
PRAD	TCGA	Validation	553	Spec
Stomach	TCGA	Validation	396	Spec
Testis	TCGA	Validation	156	Spec
Normal Blood	GSE40279	Validation	656	Spec
Normal Blood	GSE52068	Validation	312	Spec

Table 40 Esophageal cancer

Disease Status	Source	Cohort	N	Detect/Spec
Esophageal Cancer	GSE52826	Discovery		6 Detect
ESCA	TCGA	Discovery		10 Detect
Non-esophageal cancer	TCGA	Discovery		190 Spec
Healthy blood	GSE40279	Discovery		10 Spec
ESCA	TCGA	Discovery		20 Spec
Esophageal Cancer	TCGA	Validation		202 Detect
Bladder Cancer	TCGA	Validation		439 Spec
Brain Cancer	TCGA	Validation		689 Spec
Breast Cancer	TCGA	Validation		891 Spec
Cervical Cancer	TCGA	Validation		312 Spec
CRC Cancer	TCGA	Validation		459 Spec
Renal Cancer	TCGA	Validation		871 Spec
HNSC	TCGA	Validation		580 Spec
HCC	TCGA	Validation		430 Spec
Lung Cancer	TCGA	Validation		919 Spec
Ovarian Cancer	TCGA	Validation		10 Spec
Pancreatic Cancer	TCGA	Validation		195 Spec
Prostate Cancer	TCGA	Validation		553 Spec
Stomach Cancer	TCGA	Validation		397 Spec
Testis	TCGA	Validation		156 Spec
Normal Blood	GSE40279	Validation		656 Spec
Normal Blood	GSE61496	Validation		312 Spec
Normal tissues	Mix	Validation		115 Spec
Normal tissues	GSE85566	Validation		189 Spec

Table 41 Bladder cancer

Disease Status	Source	Cohort	N	Detect/Spec
Bladder Cancer	TCGA	Discovery	10	Detect
Normal bladder	GSE52955	Discovery	5	Detect
Bladder Cancer	TCGA	Discovery	10	Spec
Non-bladder cancers	TCGA	Discovery	180	Spec
Bladder Cancer	TCGA	Validation	439	Detect
Brain Cancer	TCGA	Validation	689	Spec
Breast Cancer	TCGA	Validation	891	Spec
Cervical Cancer	TCGA	Validation	312	Spec
CRC	TCGA	Validation	459	Spec
ESCA	TCGA	Validation	202	Spec
HCC	TCGA	Validation	430	Spec
Kidney Cancer	TCGA	Validation	871	Spec
Lung Cancer	TCGA	Validation	919	Spec
Ovarian Cancer	TCGA	Validation	10	Spec
PANC	TCGA	Validation	195	Spec
PRAD	TCGA	Validation	553	Spec
Stomach Cancer	TCGA	Validation	396	Spec
Testis Cancer	TCGA	Validation	156	Spec
Normal tissues	GSE50192	Validation	70	Spec
Normal Blood	GSE40279	Validation	656	Spec
Normal Blood	GSE61496	Validation	312	Spec

Table 42 Kidney cancer

Disease Status	Source	Cohort	N	Detect/Spec
Control	GSE52955	Discovery	6	Spec-Detect
Kidney Cancer	TCGA	Discovery	30	Spec-Detect
Healthy blood	GSE40279	Discovery	10	Spec-Detect
Non-kidney can	TCGA	Discovery	180	Spec-Detect
Kidney	TCGA	Validation	871	Spec-Detect
Bladder	TCGA	Validation	439	Spec-Detect
Brain	TCGA	Validation	689	Spec-Detect
Breast	TCGA	Validation	891	Spec-Detect
Cervix	TCGA	Validation	312	Spec-Detect
CRC	TCGA	Validation	459	Spec-Detect
ESCA	TCGA	Validation	202	Spec-Detect
HNSC	TCGA	Validation	580	Spec-Detect
HCC	TCGA	Validation	430	Spec-Detect
Lung	TCGA	Validation	919	Spec-Detect
Ovarian	TCGA	Validation	10	Spec-Detect
PANC	TCGA	Validation	195	Spec-Detect
PRAD	TCGA	Validation	553	Spec-Detect
Stomach	TCGA	Validation	397	Spec-Detect
Testis	TCGA	Validation	156	Spec-Detect
Normal Blood	GSE40279	Validation	656	Spec-Detect
Normal Blood	GSE61496	Validation	312	Spec-Detect
Normal tissues	Mix	Validation	189	Spec-Detect
Normal tissues	GSE85566	Validation	115	Spec-Detect

Table 33 Testicular Cancer

Disease/Status	Source	Cohort	N	Detect/Spec
Testicular Cancer	TCGA	Discovery	10	Spec-Detect
Healthy Blood	GSE40279	Discovery	10	Spec-Detect
Non-testis Cancers	TCGA	Discovery	170	Spec-Detect
Testis Cancer	TCGA	Validation	156	Spec-Detect
Bladder Cancer	TCGA	Validation	439	Spec-Detect
Brain Cancer	TCGA	Validation	689	Spec-Detect
Breast Cancer	TCGA	Validation	891	Spec-Detect
Cervical Cancer	TCGA	Validation	312	Spec-Detect
CRC	TCGA	Validation	459	Spec-Detect
ESCA	TCGA	Validation	202	Spec-Detect
HCC	TCGA	Validation	430	Spec-Detect
Kidney Cancer	TCGA	Validation	871	Spec-Detect
Lung Cancer	TCGA	Validation	919	Spec-Detect
Ovarian Cancer	TCGA	Validation	10	Spec-Detect
PANC	TCGA	Validation	195	Spec-Detect
PRAD	TCGA	Validation	553	Spec-Detect
Stomach Cancer	TCGA	Validation	396	Spec-Detect
Normal Tissues	GSE50192	Validation	70	Spec-Detect
Normal Tissues	Mix	Validation	119	Spec-Detect
Normal Blood	GSE40279	Validation	656	Spec-Detect
Normal Blood	GSE61496	Validation	312	Spec-Detect

Table 34 Pan-cancer

Disease/Status	Source	Cohort	N	Detect/Spec
Healthy Blood	GSE40279	Discovery	10	Detect
TCGA Cancers	TCGA	Discovery	170	Detect
Bladder Cancer	TCGA	Validation	439	Detect
Brain Cancer	TCGA	Validation	689	Detect
Breast Cancer	TCGA	Validation	891	Detect
Cervical Cancer	TCGA	Validation	312	Detect
CRC Cancer	TCGA	Validation	459	Detect
Esophagus Cancer	TCGA	Validation	202	Detect
HNSC Cancer	TCGA	Validation	580	Detect
HCC Cancer	TCGA	Validation	430	Detect
Lung Cancer	TCGA	Validation	919	Detect
Ovarian Cancer	TCGA	Validation	10	Detect
Pancreatic Cancer	TCGA	Validation	195	Detect
Prostate Cancer	TCGA	Validation	553	Detect
Stomach Cancer	TCGA	Validation	397	Detect
Normal Blood	GSE40279	Validation	656	Detect
Normal Blood	GSE61496	Validation	312	Detect

Table 35 Melanoma

Disease/Status	Source	Cohort	N	Detect/Spec
Bladder Cancer	TCGA	Validation	439	Detect/Spec
Brain Cancer	TCGA	Validation	689	Detect/Spec
Breast Cancer	TCGA	Validation	891	Detect/Spec
Cervical Cancer	TCGA	Validation	312	Detect/Spec
CRC Cancer	TCGA	Validation	459	Detect/Spec
Esophagus Cancer	TCGA	Validation	202	Detect/Spec
HNSC Cancer	TCGA	Validation	580	Detect/Spec
HCC Cancer	TCGA	Validation	430	Detect/Spec

Table 463 AML				
Disease	State	Source	Cohort	N
Normal	Blood	GSE40279	Validation	656
Normal	Blood	GSE61496	Validation	312
AML		GSE86409	Validation	79
AML		TCGA	Validation	140

BCD method

- The following are the steps of the Binary Categorical Differentiation method (BCD) used in the embodiments to discover polygenic DNA methylation markers for early prediction of different cancers.

We filtered the 28,754 CGIDs that are robustly unmethylated in normal tissues.

- For the discovery cohort we delineated within the list of 28,754 CG IDs that are robustly unmethylated in normal tissues, CGIDs that are categorically methylated in a particular cancer and are not methylated in unaffected tissue and normal tissues using the COUNTIF and IF functions in Microsoft excel.

$NmcCGID_x = COUNTIF(\text{beta}CGID_x\text{Cancer } n_1:n_i, ">0.2")$

$NmnCGID_x = COUNTIF(\text{beta}CGID_x\text{Normal } n_1:n_i, ">0.1")$

$DMCGID_x = IF((AND(NmcCGID_x > 0, NmnCGID_x = 0)), "TRUE", "FALSE")$

- DM CGID_x were sorted from highest to lowest number

Up to 20 top TRUE DM CGID_x positions were selected

$NmcCGID_x$ = Number of cancer patients with methylated CGID_x

Nmn = number of normal adjacent or similar tissue samples with methylated CGID_x

$\text{beta}CGID_x$ = level of methylation of CGID_x

- n = patient from 1 to i

DM = differentially methylated CGID_x

The inventors noticed that testicular and kidney cancer exhibit pervasive lack of methylation at CG IDs that are highly methylated in all tissues. We therefore used a

modification of the BCD method to discover categorically differentially methylated CG ID positions for testicular and kidney (renal) cancer which we term “BCDhypo”; categorically unmethylated in cancer and methylated in normal tissue. The following steps were used for discovery of Differentially hypomethylated CGID positions in testicular and kidney cancer.

For the discovery cohort we delineated hypomethylated CGIDs in testis or kidney that are fully methylated in normal tissues using the COUNTIF and IF functions in excel.

NucCGIDx=COUNTIF (betaCGIDxCancer n₁:n_i, “<0.2”)

NunCGIDx=COUNTIF (betaCGIDxNormal n₁:n_i, “<0.9”)

10 DHMCGIDx= IF((AND(NucCGIDx>0, NunCGIDx=0)),”TRUE”,”FALSE”)

DHM CGID positions were sorted from highest to lowest number

20 top TRUE DHM sites were selected and subjected to penalized regression analysis

NucCGIDx=Number of cancer patients with unmethylated CGID X

NunCGIDx=number of normal tissue samples with unmethylated CGID X

15 n=patient from 1 to i

DHM=differentially hypomethylated CGID

The inventors then performed on the top 20 DM (or DHM) CGIDx a penalized regression using the package penalized in R to delineate the minimal combination of

20 CGIDx that predicts cancer at highest sensitivity and specificity. The polygenic

combination of CGIDx was further tested in a multivariate linear regression equation to determine the regression coefficient between methylation levels of these CGIDs in the polygenic combination and cancer. The model was used to calculate a methylation score for each patient for a typical cancer.

25 $Ms = \alpha + \sum_{i=1}^n \beta_i CG_i$

Ms=methylation score, α =intercept, β_i =coefficient for CG ID_i, CG_i=methylation level per CG in combination. 1 to i=number of CGs in combination.

Embodiment 3. Discovery of a polygenic DNA methylation marker for liver cancer (HCC).

30

The inventors used normalized Illumina 450K DNA methylation data from GSE61258 (normal liver) and 66 randomly selected samples from the TCGA HCC collection of HCC DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDx that were discovered in
5 embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect HCC with high sensitivity and specificity in the training cohort (FIG. 5B, Table 1) (detect). A weighted DNA methylation score and a threshold value for cancer was
10 developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 8 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between HCC and other tumors (FIG. 5C, Table 2) (spec). A weighted DNA methylation score was developed for the CGIDs as
15 described in embodiment 2.

Embodiment 4. Utility of HCC polygenic DNA methylation markers for detecting HCC.

The inventors then demonstrated that the weighted HCC DNA methylation score detected HCC in a “validation cohort” that included normalized Illumina 450K DNA
20 methylation beta values from GSE76269 for 227 HCC patients for the CGIDs in Table 1. Using this method, 95% of the HCC samples were detected as HCC (FIG. 6C). A ROC curve presented in FIG. 6A reveals the specificity (1) and sensitivity (0.96) of this methylation score for detecting cancer. The inventors then demonstrated the utility of the combined spec and detect DNA methylation scores for detecting HCC and differentiating
25 between HCC and other cancers using a “validation cohort” with methylation data from GSE75041 and TCGA for HCC and 8 other types of cancer. A ROC curve presented in FIG. 7B reveals the specificity (0.97) and sensitivity (0.95) of this methylation score for differentiating HCC from other normal tissue and other cancers. These DNA methylation markers and the calculated methylation score could be used for screening and early
30 detection of cancer in people at risk as well as the general healthy population using different biomaterial from people such as tissue, feces, saliva, plasma and urine.

Embodiment 5. Discovery of a polygenic DNA methylation marker for lung cancer.

The inventors used normalized Illumina 450K DNA methylation data for 10 people from GSE61258 (normal lung) and 10 randomly selected samples from the TCGA lung cancer collection of lung cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect lung cancer with high sensitivity and specificity (Samples included both adenocarcinoma and squamous cell carcinoma) in the training cohort (FIG. 8B, Table 3) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 8 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between lung cancer and other tumors (FIG. 8C, Table 4) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

Embodiment 6. Utility of lung polygenic DNA methylation markers for detecting lung cancer.

The inventors then demonstrated that the weighted lung cancer DNA methylation score and threshold developed in embodiment 3 (detect) detects lung cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from GSE66836, GSE63704, GSE76269 and 919 lung cancer patients from TCGA. Using this method 96% of the lung cancer samples were detected as lung cancer (FIG. 9A). The inventors then demonstrated the utility of the combined spec and detect DNA methylation scores for differentiating between lung cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for lung cancer and 8 other types of cancer (FIG. 9A). A ROC curve presented in FIG. 9B reveals the specificity (0.96) and sensitivity (0.84) of this methylation score for detecting lung cancer

from other normal tissue and other cancers (FIG. 9C). These DNA methylation markers and the calculated methylation score could be used for early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

5

Embodiment 7. Discovery of a polygenic DNA methylation marker for prostate cancer.

The inventors used normalized Illumina 450K DNA methylation data for 5 people from GSE52955 (normal prostate) and 10 randomly selected samples from the TCGA prostate cancer collection of prostate cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect prostate cancer with high sensitivity and specificity in the training cohort (FIG. 10b, Table 5) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 8 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between prostate cancer and other tumors (FIG. 10C, Table 6) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

Embodiment 8. Utility of prostate cancer polygenic DNA methylation markers for detecting prostate cancer.

The inventors then demonstrated that the weighted prostate cancer DNA methylation score and threshold developed in embodiment 3 (detect) detects prostate cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from GSE73549, GSE2955, and 430 prostate cancer patients from TCGA. Using this method 99% of the prostate cancer samples were detected as prostate cancer (FIG. 11A). The inventors then demonstrated the utility of the combined spec and detect DNA methylation scores for differentiating between prostate cancer and

other cancers using a “validation cohort” with methylation data from GSE and TCGA for prostate cancer and 8 other types of cancer (FIG. 11A). A ROC curve presented in FIG. 11B reveals the specificity (0.99) and sensitivity (0.98) of this methylation score for detecting prostate cancer from other normal tissue and other cancers (FIG. 11C). These DNA methylation markers and the calculated methylation score could be used for early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 9. Discovery of a polygenic DNA methylation marker for breast cancer.

The inventors used normalized Illumina 450K DNA methylation data for 17 people from GSE60185(normal breast) and 10 randomly selected samples from the TCGA breast cancer collection of breast cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect breast cancer with high sensitivity and specificity in the training cohort (FIG. 12B, Table 7) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 8 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between breast cancer and other tumors (FIG. 12C, Table 8) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

Embodiment 10. Utility of breast cancer polygenic DNA methylation markers for detecting breast cancer.

The inventors then demonstrated that the weighted breast cancer DNA methylation score and threshold developed in embodiment 9 (detect) detects breast cancer in a “validation cohort” that included 891 breast cancer patients using normalized Illumina 450K DNA methylation beta values from GSE60185, GSE75067, and from

TCGA. Using this method 91% of the breast cancer samples were detected as breast cancer (FIG. 13A) and both DCIS and invasive cancers were detected. The inventors then demonstrated the utility of the combined spec and detect DNA methylation scores for differentiating between breast cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for breast cancer and 8 other types of cancer (FIG. 14A). A ROC curve presented in FIG. 14B reveals the specificity (0.89) and sensitivity (0.87) of this methylation score for differentiating breast cancer from other normal tissue and other cancers (FIG. 14C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of breast cancer in women at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 11. Discovery of a polygenic DNA methylation marker for colorectal cancer (CRC).

The inventors used normalized Illumina 450K DNA methylation data for 25 people from GSE(32146) (normal) and 50 randomly selected samples from the TCGA colorectal cancer collection of colorectal cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect colorectal cancer with high sensitivity and specificity in the training cohort (FIG. 15B, Table 9) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 8 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between colorectal cancer and other tumors (FIG. 15C, Table 10) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

Embodiment 12. Utility of colorectal cancer polygenic DNA methylation markers for detecting colorectal cancer.

The inventors then demonstrated that the weighted colorectal cancer DNA methylation score and threshold developed in embodiment 11 (detect) detects colorectal cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from GSE69550 and 459 colorectal cancer patients from TCGA. Using this method 98% of the colorectal cancer samples were detected as colorectal cancer (FIG. 16A). The inventors then demonstrated the utility of the combined spec and detect DNA methylation scores for differentiating between colorectal cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for colorectal cancer and 8 other types of cancer (FIG. 16A). A ROC curve presented in FIG. 16B reveals the specificity (0.96) and sensitivity (0.98) of this methylation score for detecting colorectal cancer from other normal tissues and other cancers (FIG. 16C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of cancer in people at risk for CRC as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 13. Discovery of a polygenic DNA methylation marker for pancreatic cancer.

The inventors used normalized Illumina 450K DNA methylation data for 12 people from GSE53051 (normal) and 20 randomly selected samples from the TCGA collection of pancreatic cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect pancreatic cancer with high sensitivity and specificity in the training cohort (FIG. 17B, Table 11) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 100 randomly selected DNA methylation samples

from TCGA representing 10 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between pancreatic cancer and other tumors (FIG. 17C, Table 12) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

5

Embodiment 14. Utility of pancreatic cancer polygenic DNA methylation markers for detecting pancreatic cancer.

The inventors then demonstrated that the weighted pancreatic cancer DNA methylation score and threshold developed in embodiment 13 (detect) detects pancreatic
10 cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 891 pancreatic cancer patients from TCGA. Using this method 86% of the pancreatic cancer samples were detected as pancreatic cancer (FIG. 18A). The inventors then demonstrated the utility of the combined spec and detect DNA methylation scores for differentiating between pancreatic cancer and other cancers using
15 a “validation cohort” with methylation data from GSE and TCGA for pancreatic cancer and 9 other types of cancer (FIG. 18A). A ROC curve presented in FIG. 18B reveals the specificity (0.93) and sensitivity (0.86) of this methylation score for detecting pancreatic cancer and differentiating it from other normal tissue and other cancers (FIG. 18C). These DNA methylation markers and the calculated methylation scores from the
20 methylation values could be used for early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 15. Discovery of a polygenic DNA methylation marker for brain cancer.

25 The inventors used normalized Illumina 450K DNA methylation data for 10 people from GSE65820 (normal) and 10 randomly selected samples from the TCGA collection of brain cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood
30 samples. The inventors then used the BCD method described in embodiment 2 to discover a set of binary categorical differentially methylated CGIDs that detect brain

cancer with high sensitivity and specificity in the training cohort (FIG. 19B, Table 13) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 110 randomly selected DNA methylation samples from TCGA
5 representing 11 different tumor types. The inventors used this training cohort and discovered that the detect CGID also differentiates between brain cancer and other tumors (FIG. 19C, Table 13) (detect-spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

- 10 Embodiment 16. Utility of brain cancer polygenic DNA methylation markers for detecting brain cancer.

The inventors then demonstrated that the weighted brain cancer DNA methylation score and threshold developed in embodiment 15 (detect) detects brain cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta
15 values from 689 brain cancer patients from TCGA, 40 patients from GSE58298 and 136 patients from GSE36278. Using this method 91-97% of the brain cancer samples were detected as brain cancer (FIG. 20A). The inventors then demonstrated the utility of the same CGIDs for differentiating between brain cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for brain cancer and 9
20 other types of cancer (FIG. 20A). A ROC curve presented in FIG. 22B reveals the specificity (1) and sensitivity (0.97) of this methylation score for detecting brain cancer from other normal tissue and other cancers (FIG. 20C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of cancer in people at risk as well as the general healthy population using
25 different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 17. Discovery of a polygenic DNA methylation marker for stomach cancer.

The inventors used normalized Illumina 450K DNA methylation data for 18 people from GSE99553 (normal) and 10 randomly selected samples from the TCGA
30 collection of stomach cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in

embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect stomach cancer with high sensitivity and specificity in the training cohort (FIG. 21B, Table 14) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 100 randomly selected DNA methylation samples from TCGA representing 11 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between stomach cancer and other tumors (FIG. 21C, Table 15) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

Embodiment 18. Utility of stomach cancer polygenic DNA methylation markers for detecting stomach cancer.

The inventors then demonstrated that the weighted stomach cancer DNA methylation score and threshold developed in embodiment 17 (detect) detects stomach cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 397 stomach cancer patients from TCGA. Using this method, 88% of the stomach cancer samples were detected as stomach cancer (FIG. 23A). The inventors then demonstrated the utility of the combined spec and detect DNA methylation scores for differentiating between stomach cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for stomach cancer and 10 other types of cancer (FIG. 23A). A ROC curve presented in FIG. 22B reveals the specificity (0.9 for) and sensitivity (0.9) of this methylation score for detecting stomach cancer from other normal tissues and other cancers (FIG. 22C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for screening and early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 19. Discovery of a polygenic DNA methylation marker for ovarian cancer.

The inventors used normalized Illumina 450K DNA methylation data for 5 people from GSE65820 (normal) and 10 randomly selected samples from the TCGA collection of ovarian cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in
5 embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect ovarian cancer with high sensitivity and specificity in the training cohort (FIG. 23B, Table 16) (detect). A weighted DNA methylation score and a threshold value for cancer
10 was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 100 randomly selected DNA methylation samples from TCGA representing 10 different tumor types and blood. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between ovarian cancer and other tumors (FIG. 2C, Table 17) (spec). A weighted DNA
15 methylation score was developed for the CGIDs as described in embodiment 2.

Embodiment 20. Utility of ovarian cancer polygenic DNA methylation markers for detecting ovarian cancer.

The inventors then demonstrated that the weighted ovarian cancer DNA
20 methylation score and threshold developed in embodiment 19 (detect) detects ovarian cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 114 ovarian cancer patients from TCGA. Using this method 86% of the ovarian cancer samples were detected as ovarian cancer (FIG. 24A). The inventors then demonstrated the utility of the spec DNA methylation scores for
25 differentiating between ovarian cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for ovarian cancer and 9 other types of cancer (FIG. 24A). A ROC curve presented in FIG. 24B reveals the specificity (0.99) and sensitivity (1) of this methylation score for differentiating ovarian cancer from other normal tissue and other cancers (FIG. 24C). These DNA methylation markers and the
30 calculated methylation score from the methylation values could be used for early

detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 21. Discovery of a polygenic DNA methylation marker for cervical cancer.

5 The inventors used normalized Illumina 450K DNA methylation data for 20 people from GSE46306 (normal) and 10 randomly selected samples from the TCGA collection of cervix cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in
10 embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect cervical cancer with high sensitivity and specificity in the training cohort (FIG. 25B, Table 18) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then
15 generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 8 different tumor types and blood. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between cervical cancer and other tumors (FIG. 25C, Table 19) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

20

Embodiment 22. Utility of cervical cancer polygenic DNA methylation markers for detecting cervical cancer.

 The inventors then demonstrated that the weighted cervical cancer DNA methylation score and threshold developed in embodiment 21 (detect) detects cervical
25 cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 313 cervical cancer patients from TCGA. Using this method 91% of the cervical cancer samples were detected as cervical cancer (FIG. 26A). The inventors then demonstrated the utility of the spec DNA methylation scores for differentiating between cervical cancer and other cancers using a “validation cohort” with
30 methylation data from GSE and TCGA for cervical cancer and 9 other types of cancer (FIG. 26A). A ROC curve presented in FIG. 26B reveals the specificity (0.9) and

sensitivity (0.9) of this methylation score for detecting cervical cancer and differentiating it from other normal tissue and other cancers (FIG. 26C). These DNA methylation markers and the calculated methylation scores from the methylation values could be used for early detection of cancer in people at risk as well as the general healthy population
5 using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 23. Discovery of a polygenic DNA methylation marker for Head and Neck Squamous Carcinoma (HNSC).

The inventors used normalized Illumina 450K DNA methylation data for 10
10 people from GSE(52068) (normal) and 10 randomly selected samples from the TCGA cancer collection of HNSC DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to
15 discover a polygenic set of binary categorical differentially methylated CGIDs that detect HNSC with high sensitivity and specificity in the training cohort (FIG. 27B, Table 20) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA
20 representing 12 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between HNSC and other tumors (FIG. 27C, Table 21) (spec).

Embodiment 24. Utility of Head and Neck Squamous Carcinoma (HNSC) polygenic
25 DNA methylation markers for detecting HNSC.

The inventors then demonstrated that the weighted HNSC DNA methylation score and threshold developed in embodiment 23 (detect) detects HNSC in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from GSE52068 and. Using this method 88%-96% of the HNSC samples were detected (FIG.
30 28A). The inventors then demonstrated the utility of the DNA methylation detect scores for differentiating between HNSC and other cancers using a “validation cohort” with

methylation data from GSE and TCGA for HNSC and 12 other types of cancer (FIG. 28A). A ROC curve presented in FIG. 28B reveals the specificity (0.86) and sensitivity (0.88) of this methylation score for differentiating HNSC from other normal tissue and other cancers (FIG. 28C). The markers also detect several other cancers (at relatively high sensitivity and is thus of limited specificity for these cancers). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

10

Embodiment 25. Discovery of a polygenic DNA methylation marker for ESOPHAGEAL CANCER.

The inventors used normalized Illumina 450K DNA methylation data for 10 people from GSE(52068) (normal) and 10 randomly selected samples from the TCGA cancer collection of esophageal cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect esophageal cancer with high sensitivity and specificity in the training cohort (FIG. 29B, Table 22) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 12 different tumor types. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between esophageal cancer and other tumors (FIG. 29C, Table 23) (spec).

25

Embodiment 26. Utility of ESOPHAGEAL CANCER polygenic DNA methylation markers for detecting ESOPHAGEAL CANCER.

30

The inventors then demonstrated that the weighted esophageal cancer DNA methylation score and threshold developed in embodiment 25 (detect) detects esophageal

cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from GSE52068 and. Using this method 88%-96% of the esophageal cancer samples were detected (FIG. 30A). The inventors then demonstrated the utility of the detect DNA methylation scores for differentiating between esophageal cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for esophageal cancer and 12 other types of cancer (FIG. 30A). A ROC curve presented in FIG. 30B reveals the specificity (0.86) and sensitivity (0.88) of this methylation score for differentiating esophageal cancer from other normal tissue and other cancers (FIG. 30C). The markers also detect several other cancers (at relatively high sensitivity and are thus of limited specificity for these cancers). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

15

Embodiment 27. Discovery of a polygenic DNA methylation marker for bladder cancer.

The inventors used normalized Illumina 450K DNA methylation data for 5 people from GSE52955 (normal) and 10 randomly selected samples from the TCGA collection of bladder cancer DNA methylation data as a “training” cohort. The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detects bladder cancer with high sensitivity and specificity in the training cohort (FIG. 31B, Table 24) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2. The inventors then generated a “training cohort” from 80 randomly selected DNA methylation samples from TCGA representing 13 different tumor types and normal blood. The inventors used this training cohort to discover a polygenic set of differentially methylated CGIDs between bladder cancer and other tumors (FIG. 31C, Table 25) (spec). A weighted DNA methylation score was developed for the CGIDs as described in embodiment 2.

30

Embodiment 28. Utility of bladder cancer polygenic DNA methylation markers for detecting bladder cancer.

The inventors then demonstrated that the weighted bladder cancer DNA methylation score and threshold developed in embodiment 27 (detect) detects bladder cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 439 bladder cancer patients from TCGA. Using this method 96% of the bladder cancer samples were detected as bladder cancer (FIG. 32B). The inventors then demonstrated the utility of the spec DNA methylation scores for differentiating between bladder cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for bladder cancer and 13 other types of cancer (FIG. 32B). A ROC curve presented in FIG. 32C reveals the specificity (0.86) and sensitivity (0.88) of this methylation score for detecting bladder cancer from other normal tissue and other cancers (FIG. 32C). There is cross detection at reasonably high rates however of stomach, pancreatic esophageal and colorectal cancer. These DNA methylation markers and the calculated methylation scores from the methylation values could be used for early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

20

Embodiment 29. Discovery of a polygenic DNA methylation marker for kidney cancer.

The inventors used normalized Illumina 450K DNA methylation data for kidney (renal) cancer from 10 people from GSE52955 (normal) and 10 randomly selected samples per cancer from 13 cancers in the TCGA dataset as a “training” cohort as well as normal tissues and blood (GSE40279, GSE 52955). The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that are robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD hypo method described in embodiment 2 to discover a polygenic set of binary categorical differentially methylated CGIDs that detect kidney cancer with high sensitivity and specificity in the training cohort and are specific for kidney cancer against other cancers “Detect-Spec” (FIG. 33B, Table 26) (detect-spec). A weighted DNA

30

methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2.

Embodiment 30. Utility of kidney cancer polygenic DNA methylation markers for
5 detecting kidney cancer.

The inventors demonstrated that the weighted kidney cancer DNA methylation score and threshold developed in embodiment 27 (“Detect-Spec”) detects kidney cancer in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 871 kidney cancer patients from TCGA and differentiate kidney cancer from
10 other cancers. Using this method 90% of the kidney cancer samples were detected as kidney cancer (FIG. 34A). The inventors then demonstrated the utility of the “Detect-Spec” DNA methylation scores for differentiating between kidney cancer and other cancers using a “validation cohort” with methylation data from GSE and TCGA for kidney cancer and 13 other types of cancer (FIG. 34A). A ROC curve presented in FIG.
15 34B reveals the specificity (0.87) and sensitivity (0.91) of this methylation score for detecting kidney cancer from other normal tissue and other cancers (FIG. 34C) (high crossover with HCC, brain and testis). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of kidney cancer in people at risk as well as the general healthy population
20 using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 31. Discovery of a polygenic DNA methylation marker for testicular cancer.

The inventors used normalized Illumina 450K DNA methylation data for testicular cancer from 10 people from GSE46306 (normal) and 10 randomly selected
25 samples per cancer from 13 cancers in the TCGA dataset as a “training” cohort as well as normal tissues and blood (GSE40279, GSE 61496). The inventors first shortlisted in the “training cohort” dataset 28754 CGIDs that were discovered in embodiment 1 as sites that robustly unmethylated across normal tissues and blood samples. The inventors then used the BCD hypo method described in embodiment 2 to discover a polygenic set of
30 binary categorical differentially methylated CGIDs that detect testicular cancer with high sensitivity and specificity in the training cohort and are specific for testicular cancer

against other cancers “Detect-Spec” (FIG. 35B, Table 27) (detect-spec). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2.

- 5 Embodiment 32. Utility of testicular cancer polygenic DNA methylation markers for detecting testis cancer.

 The inventors then demonstrated that the weighted testicular cancer DNA methylation score and threshold developed in embodiment 31 (“Detect-Spec”) detects testicular cancer in a “validation cohort” that included normalized Illumina 450K DNA
10 methylation beta values from 156 testicular cancer patients from TCGA and differentiate testicular cancer from other cancer. Using this method 96% of the testicular cancer samples were detected as testicular cancer (FIG. 36A). The inventors then demonstrated the utility of the “Detect-Spec” DNA methylation scores for differentiating between testicular cancer and other cancers using a “validation cohort” with methylation data from
15 GSE and TCGA for testicular cancer and 13 other types of cancer (FIG. 36A). A ROC curve presented in FIG. 36B reveals the specificity (0.97) and sensitivity (0.96) of this methylation score for detecting testicular cancer from other normal tissue and other cancers (FIG. 36C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of cancer in people at
20 risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 33. Discovery of a polygenic pan-cancer DNA methylation marker for 13 common solid tumors.

- 25 The inventors used normalized Illumina 450K DNA methylation data for 10 randomly selected samples per cancer from 13 cancers (bladder cancer, brain cancer, breast cancer, cervical cancer, colorectal cancer, esophageal cancer, HNSC, liver cancer, lung cancer, ovarian cancer, pancreatic cancer, prostate cancer, stomach cancer) in the TCGA dataset as a “training” cohort as well as normal tissues and blood from TCGA and
30 GEO. The inventors then performed a penalized regression on the combined list of CGIDs for detection of 10 different cancer listed in tables x-y and shortlisted CGIDs that

detect any of the 10 common cancers at high sensitivity and specificity (FIG. 37B, Table 28) (detect). A weighted DNA methylation score and a threshold value for cancer was developed for the CGIDs as described in embodiment 2.

- 5 Embodiment 34. Utility of pan-cancer polygenic DNA methylation markers for detecting cancer.

 The inventors then demonstrated that the weighted cancer DNA methylation score and threshold developed in embodiment 33 (“Detect”) detects 13 common cancers (bladder cancer, brain cancer, breast cancer, cervical cancer, colorectal cancer,
10 esophageal cancer, HNSC, liver cancer, lung cancer, ovarian cancer, pancreatic cancer, prostate cancer, stomach cancer) in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 3644 cancer patients from TCGA from other normal tissues. Using this method 90-95% of cancer samples were detected (FIG. 38A). A ROC curve presented in FIG. 38B reveals the specificity (0.99) and
15 sensitivity (0.95) of this methylation score for detecting 13 cancers from other normal tissue (FIG. 38C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of cancer in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

20

- Embodiment 35. Discovery of a polygenic DNA methylation marker for detection of melanoma.

 The inventors used normalized Illumina 450K DNA methylation data for 10 randomly selected melanoma samples and 220 samples from other cancers (bladder
25 cancer, brain cancer, breast cancer, cervical cancer, colorectal cancer, esophageal cancer, HNSC, liver cancer, lung cancer, ovarian cancer, pancreatic cancer, prostate cancer, stomach cancer) and normal blood in the TCGA and GEO datasets as a “training” cohort. The inventors then performed a penalized regression on the combined list of CGIDs for detection of melanoma and shortlisted CGIDs that detect melanoma at high sensitivity
30 and specificity (FIG. 39, Table 28) (detect-spec). A weighted DNA methylation score and

a threshold value for melanoma was developed for the CGIDs as described in embodiment 2.

Embodiment 36. Utility of melanoma polygenic DNA methylation marker for detecting melanoma.

The inventors then demonstrated that the weighted melanoma DNA methylation score and threshold developed in embodiment 35 (“Detect-spec”) detects melanoma in a “validation cohort” that included normalized Illumina 450K DNA methylation beta values from 475 melanoma patients from TCGA from other cancer and normal tissues. Using this method 98% of melanoma samples were detected (FIG. 40A). A ROC curve presented in FIG. 40B reveals the specificity (0.98) and sensitivity (0.95) of this methylation score for detecting melanoma from other normal tissue and other cancers (FIG. 40C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of melanoma in people at risk as well as the general healthy population using different biomaterial from the patient from tissues, feces, saliva, plasma and urine.

Embodiment 37. Discovery of a polygenic DNA methylation marker for detection of acute myeloid leukemia (AML).

The inventors used normalized Illumina 450K DNA methylation data for 10 randomly selected AML samples and 10 normal blood samples in the GEO datasets as a “training” cohort. The inventors then performed a penalized regression on the combined list of CGIDs for detection of AML and shortlisted CGIDs that detect melanoma at high sensitivity and specificity (FIG. 41, Table 27) (detect-spec). A weighted DNA methylation score and a threshold value for melanoma was developed for the CGIDs as described in embodiment 2.

Embodiment 38. Utility of acute myeloid leukemia (AML) polygenic DNA methylation markers for detecting AML in blood DNA.

The inventors then demonstrated that the weighted melanoma DNA methylation score and threshold developed in embodiment 37 (“Detect-spec”) detects AML in a

“validation cohort” that included normalized Illumina 450K DNA methylation beta values from 79 AML patients from GEO and 140 patients from TCGA and normal blood. Using this method 100% of AML samples were detected (FIG. 42A). A ROC curve presented in FIG. 42B reveals the specificity (1) and sensitivity (1) of this methylation score for detecting AML from blood (FIG. 42C). These DNA methylation markers and the calculated methylation score from the methylation values could be used for early detection of AML in people at risk as well as the general healthy population using blood DNA.

10

Embodiment 39. Bisulfite conversion, Multiplex amplification and next generation sequencing and calculation of a methylation score for predicting prostate cancer.

Blood was collected in 9-ml tubes containing K3-EDTA and processed within 1 h. Fresh blood samples were centrifuged at 1000 g for 10 min at 4 °C. The

15

supernatant was carefully transferred to a Falcon tube without disturbing the cellular layer and centrifuged again for 10 min for complete removal of any residual cells and frozen at -80C. Plasma samples are thawed, and DNA is extracted by several available methods and commercially available kits for plasma DNA extraction such as Qiagen kit for plasma DNA or EZ DNA direct extraction method. DNA is purified using

20

commercially available methods such as on AMPure XP magnetic beads and purified DNA is treated with sodium bisulfite using for example the EZ DNA bisulfite treatment kit. A library of targeted sequences is generated by two step PCR reactions (FIG. 40).

The first PCR reaction targets the specific CGIDs from table 5 and 6, note PCR1 primer has complementary sequence to second PCR2 primers (FIG. 40). The inventors used

25

human bisulfite converted genomic DNA from HEK293 cells to amplify concurrently three sequences of DNA that contain CGIDs that detect prostate cancer from HIF3A (232 base pairs region), TPM4 (213 base pairs region), and CTTN (199 base pairs region) in a multiplexed PCR reaction using the following primers in a standard Taq polymerase reaction: For CGID cg02879662; forward primer:

30

5'ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGGTAGGAGTTTTGGG

AATTGG3' and reverse primer:

5'gTgACTggAgTTCAgACgTgTgCTCTTCCgATCTCCACCCCTACAATCCCTAA3'

For CGID cg16232979; forward primer:

5'ACACTCTTTCCCTACACgACgCTCTTCCgATCT

5 NNNNNYGGTTTYGGGTTTYGTATT3'

and reverse primer:

5'gTgACTggAgTTCAgACgTgTgCTCTTCCgATCTACRCAAAAATATAAATCRACR
ATC3'

For CGID: cg14041701 and cg14498227; forward primer:

10 5'ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGTTTTGYGTTTYGGA
TTTGGGTT3'

and reverse primer:

5'gTgACTggAgTTCAgToACgTgTgCTCTTCCgATCTCATAAACAACACCTTTAAA
TAAACACTAAA3'. The amplified fragments were fractionated on an agarose gel

15

To barcode the samples, we use a second PCR reaction with the following primers:

Forward primer:

5'AATgATACggCgACCACgAgATCTACACTCTTTCCCTACACgAC3'

Barcoding primer (reverse):

20 5' CAAgCAGAAgACggCATACgAgATAGTCATCG gTgACTggAgTTCAgACgTg3'

(bold bases are the index; 200 variations of this index are used. The second set of
primers introduces the index for each patient as well as the reverse and forward
sequencing primers. A multiplex PCR1 reaction for the three markers of prostate cancer
HIF3A 232 bp, TPM4 213 bp, and CTTN 199 bp is shown on the right panel using
25 varying primer concentrations as indicated in FIG. 41.

Embodiment 40: Utility of the method of Bisulfite conversion, Multiplex amplification
and next generation sequencing and calculation of a methylation score for predicting
cancer.

30 The inventors demonstrate that embodiment 35 could be used for high throughput
prediction of prostate cancer and other cancers using plasma samples from hundreds of

patients at the same time. Indexed amplification of highly predictive CG IDs and a streamlined method for calculating a methylation score that is indicative of cancer could be used for early detection of prostate cancer and any other cancer.

- 5 Embodiment 41. Demonstration that the biomarkers selected exhibit true BCD properties are totally hypomethylated in plasma from healthy people.
- Plasma DNA was extracted from plasma prepared from 40 healthy individuals and was subjected to targeted amplification with cancer specific primers for the following cancers: liver, prostate, lung (Fig. 43) and stomach, pan cancer and CRC (Fig 44) followed by
- 10 barcoding using a second set of amplifications (PCR 2) and next generation sequencing as described in embodiments 39 and 40. All CGs exhibited very low levels of methylation in plasma from healthy people (Fig. 43 and 44).

Embodiment 42. Bioinformatics work-flow for determining DNA methylation levels.

- 15 PCR2 products are combined quantified and purified and subjected to next generation sequencing on a Miseq Illumina sequencer. Sequence is demultiplexed using Illumina software for indexed sequencing, FASTQ files are generated for each patient. A Perl text editing script <https://www.activestate.com/activeperl/downloads> is used to count Ts and Cs in the FASTQ files per patient per CG IDS and quantify the fraction of
- 20 methylated Cs in a CG ID in a patient by dividing the number of C/C+T. (see scheme in FIG. 42). The output CSV file is used for calculating the methylation score (Ms) for each patient using the equation: $Ms = \alpha + \sum_{i=1}^n \beta_i CG_i$ when α =intercept β =coefficient for CG ID i, CG=methylation level per CG in the combination of CG IDs from 1 to n. n=number of CGs in combination. MS=methylation score.

25

Applications of the inventive subject matter

The applications of the inventive subject matter are in the field of molecular diagnostics and early prediction of cancer in general. Any person skilled in the art could use the present inventive subject matter to derive similar noninvasive biomarkers for early

30 prediction of other cancers and other diseases that are accompanied by cell death and shedding of cell free DNA into the system such as neurological diseases, diabetes, heart

disease such as cirrhosis and damage to heart tissue in cardiovascular disease. The present inventive subject matter provides a path to finding exquisite methylation markers of specific cell types and tissues using the BCD and BCDhypo method. Also disclosed are methods and biomarkers for early prediction of a wide range of cancers that could be
5 used by anyone skilled in the art to detect cancer early and dramatically enhance survival rates and cure from cancer. The methods disclosed by the present invention could be used by anyone skilled in the art for routine yearly screening of healthy populations, to identify people who are starting to develop cancer and to treat them immediately and prevent the dire personal social and economic consequences of cancer mortality and
10 morbidity, as well as for monitoring “high risk” people and monitoring response to therapy in patients undergoing treatment to detect recurrence or metastasis. Adoption of the present invention described here for routine health care management by health providers and health check-up facilities will have a vast impact on reducing the burden of cancer as well as health care costs.

15 The fact that the inventive subject matter includes a number of different dependent claims does not mean that one cannot use a combination of these claims for predicting cancer. The embodiments disclosed herein for measuring and statistically analyzing and predicting cancer should not be considered limiting. Various other modifications will be apparent to those skilled in the art to measure DNA methylation in cancer patients such
20 as Illumina EPIC arrays, capture array sequencing, next generation sequencing, methylation specific PCR, epityper, restriction enzyme-based analyses and other methods found in the public domain. Similarly, there are numerous statistical methods in the public domain in addition to those listed here to use the inventive subject matter for prediction of cancer in patient samples.

25 Although the inventive subject matter has been explained in relation to its embodiments including one or more preferred embodiments, it is to be understood that many other possible modifications and variations can be made without departing from the spirit and scope of the claimed subject matter.

What is claimed is:

1. A method of using a “binary-categorical” DNA methylation signature of cancer to detect cancer comprising the step of deriving said “binary” DNA methylation signature by using a “binary-categorical differentiation (BCD)” method from genome wide DNA methylation maps.
2. The method of claim 1 wherein said genome wide DNA methylation maps are one or more of cancer cells, normal tissues and blood DNA.
3. The method of claim 1 wherein said “binary-categorical differentiation (BCD)” method includes use of one or more of the following: Illumina 27K, 450K or EPIC arrays, genome wide bisulfite sequencing, methylated DNA Immunoprecipitation (MeDIP) sequencing and hybridization with oligonucleotide arrays.
4. The method of claim 1 for detecting **hepatocellular carcinoma (HCC) Liver cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA wherein the tables and subsets are the following:

Liver_detect					
cg00370303	cg10900437	cg02012576	cg16460359	cg04035559	cg17419241
cg00931619	cg11223367	cg03768777	cg16977570	cg04085025	cg18607529
cg05040544	cg19289599	cg06233293	cg24804544	cg26523670	cg09992116
cg05739190	cg24599205				

Liver_spec					
cg12137206	cg14126493	cg06105778	cg22076972	cg13341720	cg03705926
cg09363194	cg07036412	cg02702614	cg10181419	cg05876864	cg11068343
cg17167468	cg15375239	cg00026222	cg17283781	cg16147221	cg26386472
cg14570307	cg06207432	cg07610192	cg03422204	cg11684022	cg23693289
cg21107197	cg04920951	cg20385508	cg25296314	cg20707679	cg26703661
cg00456086	cg05009389	cg19388016	cg08460435	cg04739306	cg04221886
cg26797073	cg04109768	cg05337743	cg00483503	cg18668780	cg10604002
cg27650175	cg05684891	cg26026416	cg00177496	cg14221460	cg16551483
cg13438961	cg24432073	cg21059834	cg23305567	cg04809136	cg21105227

Subset for detect:cg02012576, cg03768777, cg24804544, cg05739190; and

Subset for spec:

cg14126493.

- 5 5. The method of claim 1 for detecting **lung cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA wherein the tables and subsets are the following:

Lung_detect					
cg04223424	cg06530490	cg08944430	cg09463882	cg11017065	cg12405785
cg16405026	cg21410080	cg23141355	cg25024074		

Lung_spec					
cg05917732	cg02682457	cg23407396	cg23601095	cg23141355	cg15386964
cg02578368	cg24631970	cg27487839	cg16405026		
cg23141355	cg06530490	cg04223424	cg25470077	cg07138603	cg23460835
cg20678442	cg15436096				

Subset for detect:

cg04223424, cg23141355; and

5

Subset for spec:

cg05917732, cg25470077

6. The method of claim 1 for detecting **prostate cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

10

Prostate_detect_spec			
cg14283569	cg18085998	cg13303553	cg17929627

15

Subset for detect_spec:

cg14283569.

7. The method of claim 1 for detecting **breast cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva,

20

urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

Breast_detect					
cg13031251	cg19595750	cg16842053	cg18565473	cg09734791	cg25694349
cg09695735	cg03637878	cg26261793	cg18800085	cg16132520	cg05617413
cg17228900	cg06945936	cg08406370	cg24427504	cg26937500	cg11297107
cg02215070	cg14140647	cg05377226	cg07070305	cg24899571	cg07844931

Breast_Spec				
cg00467244	cg04194674	cg06879394	cg10720997	cg19966212
cg00722320	cg04751811	cg06998282	cg11498607	cg24525457
cg01308827	cg06282270	cg08066035	cg14862207	cg26228266
cg03113878	cg06405186	cg08296680	cg17945153	cg20180843

5

Subset for detect:

cg13031251, cg09734791, cg09695735, cg03637878; and

Subset for spec:

cg03113878, cg20180843.

10

8. The method of claim 1 for detecting **colorectal cancer** (CRC) (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

15

CRC_detect_spec					
cg08808128	cg03788131	cg09854653	cg21627760	cg08169901	cg07494047
cg01566242	cg13788592	cg24102266	cg17716617	cg16733654	

Subset for detect-spec:

9. cg09854653, cg01566242. The method of claim 1 for detecting **pancreatic cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are:

Pancreas_detect					
cg11017065	cg23833588	cg25024074	cg21277995	cg22286978	cg06241792
cg11591516	cg10096177	cg12035092	cg03611007	cg17996329	cg13807970
cg16678602	cg15386964	cg01423964	cg07900968	cg19118812	cg06728579
cg16232979	cg08406370				
Pancreas_spec					
cg01237565	cg08182975	cg15323936	cg19102272	cg01311909	cg26927232
cg22290704	cg15832577	cg20983577	cg26466027	cg09796911	cg15850155
cg15510118	cg25591377	cg16165258	cg25595541	cg18151519	cg19749445
cg14870128	cg13441142				

Subset for detect:

cg25024074, cg15386964, cg16232979; and

Subset for spec:

- 15 cg01237565, cg08182975, cg20983577, cg25591377.
10. The method of claim 1 for detecting **brain cancer (glioblastoma)** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating

from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

Brain_detect-spec					
cg06887260	cg19116006	cg10938374	cg03663746	cg25568243	cg27449131
cg24917627	cg04134305	cg09797645	cg02892595	cg13231951	cg26269703
cg09183941	cg16842053	cg26551897	cg26988692	cg07849581	cg25026703
cg06798642	cg22963915	cg04245373	cg25533556	cg27659841	cg24917627
cg04692993	cg06045408	cg06887260	cg11345323	cg17167076	cg17526812
cg19929355	cg22513169	cg22865720			

5 Subset for detect and spec:

cg19929355.

11. The method of claim 1 for detecting **gastric (stomach) cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

Stomach_detect					
cg04125371	cg05377226	cg05611779	cg06241792	cg07900968	cg09734791
cg11017065	cg12510981	cg13807970	cg15760257	cg18323466	cg19118812
cg19419279	cg19769760	cg20334243	cg26261793		
Stomach_spec					
cg17187167	cg07350470	cg05000488	cg25612145	cg09851120	cg00904726
cg04911739	cg18192294	cg11861709	cg19452853	cg02706110	cg03768513
cg05611779	cg06981182	cg06118999	cg04812509	cg10131095	cg05339066
cg19235339					

15

Subset for detect:

cg05611779, cg09734791, cg15760257; and

Subset for spec:

cg05611779, cg19235339.

12. The method of claim 1 for detecting **ovarian cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

Ovarian_detect					
cg24339193	cg04008429	cg06849719	cg22694153	cg11252337	cg15804105
cg12479674	cg06961071	cg21210985	cg01556502	cg02537149	cg23983315
cg03597143	cg27209395				
Ovarian_spec					
cg00895834	cg01159194	cg01961086	cg02649698	cg03345116	cg03456771
cg09392827	cg10178270	cg10581012	cg13459217	cg15701612	cg17130982
cg05901462	cg07068768	cg08389588	cg09173621	cg19276014	cg19846609
cg18476766	cg19129687				

Subset for detect:

cg24339193, cg22694153, cg11252337, cg21210985; and

Subset for spec:

cg07068768, cg19846609.

13. The method of claim 1 for detecting **cervical cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

Cervix_detect					
cg00578154	cg13644629	cg01423964	cg15745619	cg04522671	cg21621906
cg00757182	cg14962363	cg01601746	cg17228900	cg07126167	cg22806837
cg08134106	cg22922289	cg09260640	cg24625128	cg11259628	cg27420520
cg08535260	cg23141355	cg09734791	cg25024074		
Cervix_spec					
cg00829990	cg02401399	cg04996219	cg07066594	cg07195011	cg07576142
cg09260640	cg12961842	cg13668618	cg18543270		

Subset for detect:

cg00757182, cg01601746; and

Subset for spec:

5 cg07066594, cg09260640, cg12961842.

14. The method of claim 1 for detecting **head and neck squamous cell carcinoma**

(detect) and for differentiating from other tumors (spec) by measuring DNA

methylation levels of one or more of the CG IDs from the following tables and

subsets and deriving a “cancer methylation score” using either a “binary

10 categorical” or linear regression equation and Receiver operating characteristics

(ROC) assays in DNA originating from the tumor in biological material derived

from the patient such as saliva, urine, feces and cell free plasma DNA, wherein

the tables and subsets are the following:

HNSC_detect					
cg01613638	cg02776314	cg03280624	cg04524120	cg05151803	cg07746323
cg07900968	cg08406370	cg11108676	cg12083965	cg15397448	cg17428324
cg18403606	cg20334243	cg26770917	cg27009208	cg27420520	
HNSC_spec					
cg01217080	cg03522799	cg06672120	cg09136346	cg10155875	cg18006328
cg18443253	cg19287220				

15 Subset for detect:

cg07900968, cg20334243, cg27420520; and

Subset for spec:

cg18006328, cg19287220.

15. The method of claim 1 for detecting **esophageal carcinoma** (detect) and for

20 differentiating from other tumors (spec) by measuring DNA methylation levels of

one or more of the CG IDs from the following tables and subsets and deriving a

“cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

Esophagus_detect					
cg03280624	cg03735888	cg06530490	cg06963053	cg08944430	cg09734791
cg11017065	cg12035092	cg18344922	cg19118812	cg20334243	cg22128431
cg23141355	cg24740531	cg27009208	cg27420520		
Esophagus_spec					
cg00532449	cg02763101	cg04743654	cg08055087	cg08932440	cg09556952
cg10608333	cg12473285	cg12966367	cg17579667	cg17949440	cg18723937
cg21554552	cg22647407	cg23286646	cg23730575	cg27569446	

Subset Subset for detect:

cg03280624, cg03735888, cg27420520, cg09734791; and

Subset for spec:

cg09556952, cg12473285.

16. The method of claim 1 for detecting **bladder cancer** (detect) and for differentiating from other tumors (spec) by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:

Bladder_detect					
cg01423964	cg04223424	cg01556502	cg23141355	cg10723962	cg09734791
cg25024074	cg21039778	cg19113641	cg04342821	cg09008417	cg05873285
Bladder_spec					
cg14773260	cg08827322	cg05085809	cg13544006	cg02293936	cg10153335
cg05039004	cg14718568	cg07790170	cg12995090	cg16495265	cg01508045
cg04933208	cg03231163	cg06708937	cg00675569	cg06312813	cg14126688
cg12984729	cg02384661	cg08307030	cg26884027	cg20540209	cg04355159
cg08857479	cg23395715	cg22006640	cg26279336	cg19898108	cg17446010
cg12911122					

Subset for detect:

cg04223424, cg10723962, cg25024074; and

Subset for spec:

cg13544006.

- 5 17. The method of claim 1 for detecting **renal (kidney) cancer** (detect) and for
differentiating from other tumors (spec) by measuring DNA methylation levels of
one or more of the CG IDs from the following tables and subsets and deriving a
“cancer methylation score” using either a “binary categorical” or linear regression
equation and Receiver operating characteristics (ROC) assays in DNA originating
10 from the tumor in biological material derived from the patient such as saliva,
urine, feces and cell free plasma DNA, wherein the tables and subsets are the
following:

Renal cancer_detect_spec					
cg08884571	cg00011225	cg14535274	cg05507490	cg10367244	cg07190763
cg02971546	cg19990785	cg03084949	cg08765317	cg02820958	cg23346709
cg26642667					

Subset for detect spec:

- 15 cg08884571, cg00011225, cg00011225.
18. The method of claim 1 for detecting **testicular cancer** (detect) and for
differentiating from other tumors (spec) by measuring DNA methylation levels of
one or more of the CG IDs from the following tables and subsets and deriving a
“cancer methylation score” using either a “binary categorical” or linear regression
equation and Receiver operating characteristics (ROC) assays in DNA originating
20 from the tumor in biological material derived from the patient such as saliva,
urine, feces and cell free plasma DNA, wherein the tables and subsets are the
following:.

Testicular_detect_spec					
cg14531093	cg11777290	cg13283877	cg03966099	cg26939375	cg26297530
cg16157016	cg14002772	cg09890687	cg05134918	cg09719850	cg06456125
cg17978367	cg20100671	cg26627956	cg19789755	cg14393609	cg21634331

cg02039634	cg25159927	cg01895439			
------------	-------------------	------------	--	--	--

Subset for detect and spec:

cg14531093, cg25159927.

19. The method of claim 1 for detecting 13 **common solid tumors (Pan cancer)** by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristics (ROC) assays in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:.

Pan-cancer_detect					
cg01423964	cg04223424	cg06530490	cg06543087	cg07900968	cg08406370
cg08848774	cg09734791	cg10723962	cg11017065	cg15759056	cg15760257
cg16405026	cg17228900	cg21277995	cg22286978	cg23141355	cg24427504
cg25024074	cg27420520				

Subset for detect:

cg10723962, cg15759056, cg24427504, cg25024074.

20. The method of claim 1 for detecting **blood cancers** by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “cancer methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristic (ROC) assay in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:.
21. The method of claim 1 for detecting **acute myeloid leukemia (AML)** by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “AML methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristic (ROC) assay in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma

DNA, wherein the tables and subsets are the following:

AML_detect-spec					
cg00594866	cg05608159	cg18658397	cg18780412	cg20439288	cg22828045
cg25375340					

Subset for detect -spec:

cg18658397, cg18780412, cg20439288, cg22828045, cg25375340.

5

22. The method of claim 1 for detecting melanoma by measuring DNA methylation levels of one or more of the CG IDs from the following tables and subsets and deriving a “melanoma methylation score” using either a “binary categorical” or linear regression equation and Receiver operating characteristic (ROC) assay in DNA originating from the tumor in biological material derived from the patient such as saliva, urine, feces and cell free plasma DNA, wherein the tables and subsets are the following:.

10

Melanoma_detect-spec					
cg00325866	cg01228636	cg04616691	cg04824711	cg05569109	cg06019853
cg08611163	cg10830649	cg12593303	cg15307891	cg18866529	cg19530886
cg19634213	cg20146030	cg21755725	cg22280705	cg24217704	cg24678095
cg26127345	cg27084903				

Subset for detect-spec:

cg15307891, cg18866529, cg27084903.

15

23. A kit for detecting cancer, comprising equipment and one or more reagents used for deriving DNA methylation measurements of the DNA methylation score according to the method of claim 1.

20

24. A kit for detecting hepatocellular carcinoma (HCC) comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 4.
- 5 25. A kit for detecting lung cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 5.
26. A kit for detecting prostate cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 6.
- 10 27. A kit for detecting breast cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 7.
28. A kit for detecting colorectal cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 8.
- 15 29. A kit for detecting pancreatic cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 9.
- 20 30. A kit for detecting brain cancer “glioblastoma” comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 10.
31. A kit for detecting gastric (stomach) cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 11.
- 25 32. A kit for detecting ovarian cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 12.
- 30 33. A kit for detecting cervical cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 13.

34. A kit for detecting head and neck squamous cell carcinoma (HNSC) comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 14.
35. A kit for detecting esophageal carcinoma comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 15.
36. A kit for detecting bladder cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 16.
37. A kit for detecting renal kidney cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 17.
38. A kit for detecting testicular cancer comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 18.
39. A kit for detecting common solid tumors comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 19.
40. A kit for detecting blood cancers comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 20.
41. A kit for detecting acute myeloid leukemia (AML) comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 21.
42. A kit for detecting melanoma comprising equipment and one or more reagents for DNA methylation measurements of the DNA methylation signature according to the method of claim 22.
43. The methods of claims 1-22 further comprising the step of using DNA pyrosequencing methylation assays for predicting cancer by using DNA methylation combination.

44. The methods of claims 1-22 further comprising the step of using multiplexed amplification targeted bisulfite sequencing on a next generation sequencer such as the Illumina MiSeq for detecting cancer.
45. The methods of claims 1-22 further comprising the step of using a
5 “binary categorical assay” for calculating a methylation score for predicting cancer.
46. The methods of claims 1-22 further comprising the step of using a multivariate linear regression equation for calculating a methylation score to predict cancer.
47. The methods of claims 1-22 further comprising the step of using a Receiver operating characteristics (ROC) assay to define a “methylation score” threshold
10 differentiating cancer from noncancer and tissue of origin by using measurements of DNA methylation combinations.
48. A method for identifying DNA methylation signature and a “polygenic DNA methylation marker” for detecting tissue of origin of cell free DNA in body fluids such as saliva, urine, plasma, feces or tissue biopsy comprising using the “binary-
15 categorical differentiation (BCD)” method on genome wide DNA methylation maps of a specific tissue and genome wide DNA methylation maps of all other normal tissues and blood DNA, such as Illumina 27K, 450K or EPIC arrays, genome wide bisulfite sequencing, methylated DNA Immunoprecipitation (MeDIP) sequencing or hybridization with oligonucleotide arrays comprising the
20 step of performing statistical analysis on DNA methylation measurements obtained from samples.
49. The method according to claim 48 for further identifying “polygenic DNA methylation markers” of a diseased tissue such as neurons in brain, pancreas in diabetes, heart in ischemia and heart disease wherein an identification such as
25 dying neurons CF DNA in plasma can be used for early detection of disease such as Alzheimer disease.
50. The method according to claim 48, wherein said DNA methylation measurements are obtained by performing Illumina Beadchip 450K or EPIC assay of DNA extracted from one or more samples.
- 30 51. The method according to claim 48, wherein said DNA methylation measurements are obtained by performing Illumina next generation sequencing on one or more

of the following platforms: iSeq, MiniSeq, MiSeq or NextSeq sequencers, torrent sequencing, DNA pyrosequencing, mass spectrometry based (Epityper™) and PCR based methylation assays of DNA extracted from sample.

- 5 52. The method according to claim 48, wherein said statistical analysis includes Pearson correlation.
53. The method according to claim 48, wherein said statistical analysis includes Receiver operating characteristics (ROC) assays.
54. The method according to claim 48, wherein said statistical analysis includes hierarchical clustering analysis assays.