METHOD AND SYSTEM FOR DETERMINING CANCER STATUS

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 62/104,785, filed January 18, 2015, and is a Continuation of U.S. Application No. 14/986,520, filed December 31, 2015, which the applications are incorporated herein by reference.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in ASCII format *via* EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on January 12, 2016 is named 49697-701.601_SL.txt and is 846 kilobytes in size.

BACKGROUND OF THE INVENTION

[0003] Cancer is a leading cause of deaths worldwide, with annual cases expected to increase from 14 million in 2012 to 22 million during the next two decades (WHO). Diagnostic procedures, in some cases, begin only after a patient is already present with symptoms, leading to costly, invasive, and time-consuming procedures. In addition, inaccessible areas sometimes prevent an accurate diagnosis. Further, high cancer morbidities and mortalities are associated with late diagnosis.

SUMMARY OF THE INVENTION

[0004] Disclosed herein, in certain embodiments, are methods, systems, platform, non-transitory computer-readable medium, services, and kits for determining a cancer type in an individual. In some embodiments, also described herein include methods, systems, platform, non-transitory computer-readable medium, services, and kits for early detection of cancer. In additional embodiments, described herein include methods, systems, platform, non-transitory computer-readable medium, services, and kits for non-invasive detection of cancer. In still additional embodiments, described herein include methods, systems, platform, non-transitory computer-readable medium, services, and kits for distinguishing different cancer stages. In other embodiments, described herein include methods, systems, platform, non-transitory computer-readable medium, services, and kits for determining the prognosis of a cancer in an individual in need thereof, prediction of a treatment response, and treatment response monitoring. In further embodiments, described herein include methods, systems, platform, non-transitory computer-readable medium, services, and kits for generating a CpG methylation profile database, and probes used in generating CpG methylation data.

[0004a] In one aspect, there is provided a computing platform for utilizing CpG cancer methylation data for generation of a cancer CpG methylation profile database, comprising:

(a) a first computing device comprising a processor, a memory module, an operating system, and a computer program including instructions executable by the processor to create a data acquisition application for generating CpG methylation data from a set of biological samples, the data acquisition application comprising:

(1) a sequencing module configured to operate a sequencing device to generate CpG methylation data from a set of biological samples, wherein the set comprises a first cancerous biological sample, a second cancerous biological sample, a third cancerous biological sample, a first normal biological sample, a second normal biological sample, and a third normal biological sample; wherein the first, second, and third cancerous biological samples are different; and wherein the first, second, and third normal biological samples are different; and

(2) a data receiving module configured to receive:

(i) a first pair of CpG methylation datasets generated from the first cancerous biological sample and the first normal biological sample, wherein CpG methylation data generated from the first cancerous biological sample form a first dataset within the first pair of datasets, CpG methylation data generated from the first normal biological sample form a second dataset within the first pair of datasets, and the first cancerous biological sample and the first normal biological sample are from the same biological sample source;

(ii) a second pair of CpG methylation datasets generated from the second normal biological sample and the third normal biological sample, wherein CpG methylation data generated from the second normal biological sample form a third dataset within the second pair of datasets, CpG methylation data generated from the third normal biological sample form a fourth dataset within the second pair of datasets, and the first, second, and third normal biological samples are different; and

(i) a third pair of CpG methylation datasets generated from the second cancerous biological sample and the third cancerous biological sample, wherein CpG methylation data generated from the second cancerous biological sample form a fifth dataset within the third pair of datasets, CpG methylation data generated from the third cancerous biological sample form a sixth dataset within the third pair of datasets, and the first, second, and third cancerous biological samples are different; and

(b) a second computing device comprising a processor, a memory module, an operating system, and a computer program including instructions executable by the processor to create a data analysis application for generating a cancer CpG methylation profile database, the data analysis application comprising a data analysis module configured to:

(1) generate a pair-wise methylation difference dataset from the first, second, and third pair of datasets; and

(2) analyze the pair-wise methylation difference dataset with a control dataset by a machine learning method to generate the cancer CpG methylation profile database, wherein

(i) the machine learning method comprises: identifying a plurality of makers and a plurality of weights based on a top score, and classifying the samples based on the plurality of markers and the plurality of weights; and

(ii) the cancer CpG methylation profile database comprises a set of CpG methylation profiles and each CpG methylation profile represents a cancer type.

[0004b] In one aspect, there is provided a computer-implemented method for generating a cancer CpG methylation profile database, comprising;

(a) generating CpG methylation data from a set of biological samples by a sequencing method, wherein the set comprises a first cancerous biological sample, a second cancerous biological sample, a third cancerous biological sample, a first normal biological sample, a second normal biological sample, and a third normal biological sample; wherein the first, second, and third cancerous biological samples are different; and wherein the first, second, and third normal biological samples are different;

(b) obtaining a first pair of CpG methylation datasets, with a first processor, generated from the first cancerous biological sample and the first normal biological sample, wherein CpG methylation data generated from the first cancerous biological sample form a first dataset within the first pair of datasets, CpG methylation data generated from the first normal biological sample form a second dataset within the first pair of datasets, and the first cancerous biological sample and the first normal biological sample are from the same biological sample source;

(c) obtaining a second pair of CpG methylation datasets, with the first computing device, generated from the second normal biological sample and the third normal biological sample, wherein CpG methylation data generated from the second normal biological sample form a third dataset within the second pair of datasets, CpG methylation data generated from the third normal biological sample form a fourth dataset within the second pair of datasets, and the first, second, and third normal biological samples are different;

(d) obtaining a third pair of CpG methylation datasets, with the first computing device, generated from the second cancerous biological sample and the third cancerous biological sample, wherein CpG methylation data generated from the second cancerous biological sample form a fifth dataset within the third pair of datasets, CpG methylation data generated from the third cancerous biological sample form a sixth dataset within the third pair of datasets, and the first, second, and third cancerous biological samples are different;

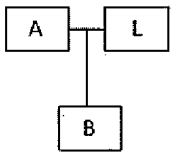
(e) generating a pair-wise methylation difference dataset, with a second processor, from the first, second, and third pair of datasets; and

(f) analyzing the pair-wise methylation difference dataset with a control dataset by a machine learning method to generate the cancer CpG methylation profile database, wherein

(1) the machine learning method comprises: identifying a plurality of makers and a plurality of weights based on a top score, and classifying the samples based on the plurality of markers and the plurality of weights; and

(2) the cancer CpG methylation profile database comprises a set of CpG methylation profiles and each CpG methylation profile represents a cancer type.

[0004c] In one aspect, there is provided a probe panel comprising a plurality of probes, each probe is the probe of Formula I



Formula I

wherein

A is a first target-binding region;

B is a second target-binding region; and

L is a linker region;

wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 30 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-1775; B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 12 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-1775; L is attached to A; and B is attached to either A or L.

WHAT IS CLAIMED IS:

1. A computing platform for utilizing CpG cancer methylation data for generation of a cancer CpG methylation profile database, comprising:

(a) a first computing device comprising a processor, a memory module, an operating system, and a computer program including instructions executable by the processor to create a data acquisition application for generating CpG methylation data from a set of biological samples, the data acquisition application comprising:

(1) a sequencing module configured to operate a sequencing device to generate CpG methylation data from a set of biological samples, wherein the set comprises a first cancerous biological sample, a second cancerous biological sample, a third cancerous biological sample, a first normal biological sample, a second normal biological sample, and a third normal biological sample; wherein the first, second, and third cancerous biological samples are different; and wherein the first, second, and third normal biological samples are different; and

(2) a data receiving module configured to receive:

(i) a first pair of CpG methylation datasets generated from the first cancerous biological sample and the first normal biological sample, wherein CpG methylation data generated from the first cancerous biological sample form a first dataset within the first pair of datasets, CpG methylation data generated from the first normal biological sample form a second dataset within the first pair of datasets, and the first cancerous biological sample and the first normal biological sample are from the same biological sample source;

(ii) a second pair of CpG methylation datasets generated from the second normal biological sample and the third normal biological sample, wherein CpG methylation data generated from the second normal biological sample form a third dataset within the second pair of datasets, CpG methylation data generated from the third normal biological sample form a fourth dataset within the second pair of datasets, and the first, second, and third normal biological samples are different; and

(iii)a third pair of CpG methylation datasets generated from the second cancerous biological sample and the third cancerous biological sample, wherein CpG methylation data generated from the second cancerous biological sample form a fifth dataset within the third pair of datasets, CpG methylation data generated from the third cancerous biological sample form a sixth dataset within the third pair of datasets, and the first, second, and third cancerous biological samples are different; and

(b) a second computing device comprising a processor, a memory module, an operating system, and a computer program including instructions executable by the processor to create a data analysis application for generating a cancer CpG methylation profile database, the data analysis application comprising a data analysis module configured to:

(1) generate a pair-wise methylation difference dataset from the first, second, and third pair of datasets; and

(2) analyze the pair-wise methylation difference dataset with a control dataset by a machine learning method to generate the cancer CpG methylation profile database, wherein

(i) the machine learning method comprises: identifying a plurality of makers and a plurality of weights based on a top score, and classifying the samples based on the plurality of markers and the plurality of weights; and

(ii) the cancer CpG methylation profile database comprises a set of CpG methylation profiles and each CpG methylation profile represents a cancer type.

2. The platform of claim 1, wherein the generating the pair-wise methylation difference dataset comprises:

(a) calculating a difference between the first dataset and the second dataset within the first pair of datasets;

(b) calculating a difference between the third dataset and the fourth dataset within the second pair of datasets; and

(c) calculating a difference between the fifth dataset and the sixth dataset within the third pair of datasets.

3. The platform of claim 1, wherein the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model.
4. The platform of claim 1, wherein the CpG methylation data is generated from an extracted genomic DNA treated with a deaminating agent.
5. The platform of claim 4, wherein the sequence device is further configured to analyze the extracted genomic DNA by a next generation sequencing method to generate the CpG methylation data.
6. The platform of claim 5, wherein the next generation sequencing method is a digital PCR sequencing method.
7. The platform of claim 1, wherein the methylation profile comprises at least 10, 20, 30, 40, 50, 100, 200, or more of biomarkers selected from the group consisting of Tables 15-18.
8. The platform of claim 1, wherein the methylation profile comprises at least 10, 20, 30, 40, 50, 100, 200, or more biomarkers selected from cg20468939, cg24790419, cg26836479, cg16911583, cg15139596, cg16927606, cg12967050, cg21122474, cg06064964, cg11779113, cg12042264, cg27377213, cg26680502, cg12504877, cg21913888, cg26683005, cg24166457, cg27141915, cg17122157, cg09844573, cg03087897, cg24706505, cg17126555, cg13911392, cg18901104, cg25982880, cg15797834, cg27125093, cg17518965, cg20695297, cg04858553, cg09419005, cg25490145, cg11252953, cg18456621, cg07058988, cg17864646, cg06153925, cg27410601, cg03297901, cg06853339, cg12900649, cg27219182, cg15759721, cg27023597, cg02782634, cg18942579, cg01409343, cg10530767, cg26112797, cg00253248, cg01722297, cg22589778, cg07137244, cg04147906, cg23878564, cg07860918, cg00206490, cg07644807, cg00558804, cg05304979, cg27598656, cg03549146, cg22190721, cg01660934, cg02358862, cg23093496, cg07641284, cg01681367, cg26769927, cg08480068, cg02914427, cg03653601, cg01990910, cg00933696, cg09866569, cg20357538, cg22460896, cg07116712, cg10186131, cg06380123, cg18610205, cg12353452, cg10590292, cg00037681, cg05596756, cg03569637, cg02522196, cg11655490, cg19693177, cg26363363, cg21249754, cg23147227, cg01657186, cg23764129, cg04514998, cg07332880, cg16061668, cg25574765,

cg14088196, cg03758697, cg05398700, cg14058476, cg18158859, cg19300307, cg18842353, cg10732611, cg24480810, cg02053964, cg25922751, cg25954028, cg14642045, cg24165921, cg18215449, cg16402452, cg21376733, cg16509569, cg08075204, cg14556909, cg07119472, cg14999168, cg09399878, cg02874908, cg10542975, cg15698795, cg11791526, cg00862408, cg16260696, cg00220455, cg20826709, cg11436362, cg13924996, cg07420137, cg24301930, cg13395086, cg20136100, cg09153080, cg09902130, cg07380416, cg27284288, cg13912307, cg10511890, cg00242035, cg04314978, cg25225070, cg20411756, cg24247537, cg04330884, cg23130731, cg04888360, cg00907272, cg05979232, cg00025044, cg04441857, cg09684112, cg27388962, cg05931497, cg13408086, cg13555415, cg22552736, cg16191087, cg13925432, cg13464240, cg14633252, cg19252956, cg00015530, cg08632810, cg12737392, cg26769700, cg03218479, cg02609337, cg10351284, cg23554164, cg19021985, cg21031128, cg19421584, cg17984956, cg05177060, cg24107852, cg25652701, cg00282244, cg18887230, cg08486903, cg09335715, cg12629796, cg16454130, cg26433975, cg10673833, cg06787669, cg12192582, cg05098343, cg07573366, cg11105292, cg05287480, cg16748008, cg16644023, cg06488150, cg09450197, cg20336172, cg08858130, cg12098228, cg26811313, cg25432518, cg16622899, cg12359001, cg01209642, cg14564351, cg23429794, cg26401541, cg20046343, cg20847580, cg03431741, cg07417146, cg09001226, cg06482498, cg03891050, cg00899907, cg13597051, cg18113826, cg04859102, cg01620360, cg14083015, cg15046123, cg03190513, cg01456691, cg17207512, cg20510285, cg01149192, cg05614346, cg06439655, cg11334870, cg08912922, cg23021796, cg24835948, cg10393744, cg07428959, cg17694130, cg03956042, cg19266387, cg13512830, cg19982684, cg22513455, cg07186032, cg08052292, cg27366280, cg06825448, cg25451702, cg08098128, cg13821008, cg27405400, cg09366118, cg15341833, cg02233149, cg14247287, cg23824762, cg01604601, cg05656900, cg08132573, cg24686918, cg05352688, cg18384097, cg16266227, cg19675731, cg21461981, cg25765104, cg26394055, cg20685713, cg23589035, cg01903374, cg23612220, cg26315985, cg18856478, cg23229016,

cg21004490, cg24742520, cg23013029, cg19704755, cg07589991, cg10055231, and cg26017930.

9. The platform of claim 1, wherein the methylation profile comprises at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, or more biomarkers selected from Table 20.

10. The platform of claim 1, wherein the cancer type is a solid cancer type or a hematologic malignant cancer type.

11. The platform of claim 1, wherein the cancer type comprises acute myeloid leukemia (LAML or AML), acute lymphoblastic leukemia (ALL), adrenocortical carcinoma (ACC), bladder urothelial cancer (BLCA), brain stem glioma, brain lower grade glioma (LGG), brain tumor, breast cancer (BRCA), bronchial tumors, Burkitt lymphoma, cancer of unknown primary site, carcinoid tumor, carcinoma of unknown primary site, central nervous system atypical teratoid/rhabdoid tumor, central nervous system embryonal tumors, cervical squamous cell carcinoma, endocervical adenocarcinoma (CESC) cancer, childhood cancers, cholangiocarcinoma (CHOL), chordoma, chronic lymphocytic leukemia, chronic myelogenous leukemia, chronic myeloproliferative disorders, colon (adenocarcinoma) cancer (COAD), colorectal cancer, craniopharyngioma, cutaneous T-cell lymphoma, endocrine pancreas islet cell tumors, endometrial cancer, ependymoblastoma, ependymoma, esophageal cancer (ESCA), esthesioneuroblastoma, Ewing sarcoma, extracranial germ cell tumor, extragonadal germ cell tumor, extrahepatic bile duct cancer, gallbladder cancer, gastric (stomach) cancer, gastrointestinal carcinoid tumor, gastrointestinal stromal cell tumor, gastrointestinal stromal tumor (GIST), gestational trophoblastic tumor, glioblstoma multiforme glioma GBM), hairy cell leukemia, head and neck cancer (HNSD), heart cancer, Hodgkin lymphoma, hypopharyngeal cancer, intraocular melanoma, islet cell tumors, Kaposi sarcoma, kidney cancer, Langerhans cell histiocytosis, laryngeal cancer, lip cancer, liver cancer, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma [DLBCL), malignant fibrous histiocytoma bone cancer, medulloblastoma, medullo epithelioma, melanoma, Merkel cell carcinoma, Merkel cell skin carcinoma, mesothelioma (MESO), metastatic squamous neck cancer with occult primary, mouth cancer, multiple endocrine neoplasia syndromes, multiple myeloma, multiple myeloma/plasma cell neoplasm, mycosis fungoides, myelodysplastic syndromes, myeloproliferative neoplasms, nasal cavity cancer, nasopharyngeal cancer,

neuroblastoma, Non-Hodgkin lymphoma, nonmelanoma skin cancer, non-small cell lung cancer, oral cancer, oral cavity cancer, oropharyngeal cancer, osteosarcoma, other brain and spinal cord tumors, ovarian cancer, ovarian epithelial cancer, ovarian germ cell tumor, ovarian low malignant potential tumor, pancreatic cancer, papillomatosis, paranasal sinus cancer, parathyroid cancer, pelvic cancer, penile cancer, pharyngeal cancer, pheochromocytoma and paraganglioma (PCPG), pineal parenchymal tumors of intermediate differentiation, pineoblastoma, pituitary tumor, plasma cell neoplasm/multiple myeloma, pleuropulmonary blastoma, primary central nervous system (CNS) lymphoma, primary hepatocellular liver cancer, prostate cancer such as prostate adenocarcinoma (PRAD), rectal cancer, renal cancer, renal cell (kidney) cancer, renal cell cancer, respiratory tract cancer, retinoblastoma, rhabdomyosarcoma, salivary gland cancer, sarcoma (SARC), Sezary syndrome, skin cutaneous melanoma (SKCM), small cell lung cancer, small intestine cancer, soft tissue sarcoma, squamous cell carcinoma, squamous neck cancer, stomach (gastric) cancer, supratentorial primitive neuroectodermal tumors, T-cell lymphoma, testicular cancer testicular germ cell tumors (TGCT), throat cancer, thymic carcinoma, thymoma (THYM), thyroid cancer (THCA), transitional cell cancer, transitional cell cancer of the renal pelvis and ureter, trophoblastic tumor, ureter cancer, urethral cancer, uterine cancer, uterine cancer, uveal melanoma (UVM), vaginal cancer, vulvar cancer, Waldenstrom macroglobulinemia, or Wilm's tumor.

12. The platform of claim 1, wherein the control dataset comprises a set of methylation profiles, wherein each said methylation profile is generated from a biological sample obtained from a known cancer type.

13. The platform of claim 1, wherein the biological samples comprise a circulating tumor DNA sample or a tissue sample.

14. A computer-implemented method for generating a cancer CpG methylation profile database, comprising:

    a. generating CpG methylation data from a set of biological samples by a sequencing method, wherein the set comprises a first cancerous biological sample, a second cancerous biological sample, a third cancerous biological sample, a first normal biological sample, a second normal biological sample, and a third normal biological

sample; wherein the first, second, and third cancerous biological samples are different; and wherein the first, second, and third normal biological samples are different;

b. obtaining a first pair of CpG methylation datasets, with a first processor, generated from the first cancerous biological sample and the first normal biological sample, wherein CpG methylation data generated from the first cancerous biological sample form a first dataset within the first pair of datasets, CpG methylation data generated from the first normal biological sample form a second dataset within the first pair of datasets, and the first cancerous biological sample and the first normal biological sample are from the same biological sample source;

c. obtaining a second pair of CpG methylation datasets, with the first computing device, generated from the second normal biological sample and the third normal biological sample, wherein CpG methylation data generated from the second normal biological sample form a third dataset within the second pair of datasets, CpG methylation data generated from the third normal biological sample form a fourth dataset within the second pair of datasets, and the first, second, and third normal biological samples are different;

d. obtaining a third pair of CpG methylation datasets, with the first computing device, generated from the second cancerous biological sample and the third cancerous biological sample, wherein CpG methylation data generated from the second cancerous biological sample form a fifth dataset within the third pair of datasets, CpG methylation data generated from the third cancerous biological sample form a sixth dataset within the third pair of datasets, and the first, second, and third cancerous biological samples are different;

e. generating a pair-wise methylation difference dataset, with a second processor, from the first, second, and third pair of datasets; and

f. analyzing the pair-wise methylation difference dataset with a control dataset by a machine learning method to generate the cancer CpG methylation profile database, wherein

(1) the machine learning method comprises: identifying a plurality of makers and a plurality of weights based on a top score, and classifying the samples based on the plurality of markers and the plurality of weights; and

(2) the cancer CpG methylation profile database comprises a set of CpG methylation profiles and each CpG methylation profile represents a cancer type.

15. The computer-implemented method of claim 14, wherein step e) further comprises
    a. calculating a difference between the first dataset and the second dataset within the first pair of datasets;
    b. calculating a difference between the third dataset and the fourth dataset within the second pair of datasets; and
    c. calculating a difference between the fifth dataset and the sixth dataset within the third pair of datasets.

16. The computer-implemented method of claim 14, wherein the machine learning method utilizes an algorithm selected from one or more of the following: a principal component analysis, a logistic regression analysis, a nearest neighbor analysis, a support vector machine, and a neural network model

17. The computer-implemented method of claim 14, wherein the CpG methylation data is generated from an extracted genomic DNA treated with a deaminating agent.

18. The computer-implemented method of claim 14, wherein the methylation profile comprises at least 10, 20, 30, 40, 50, 100, 200, or more of biomarkers selected from the group consisting of Tables 15-18.

19. The computer-implemented method of claim 14, wherein the methylation profile comprises at least 10, 20, 30, 40, 50, 100, 200, or more biomarkers selected from cg20468939, cg24790419, cg26836479, cg16911583, cg15139596, cg16927606, cg12967050, cg21122474, cg06064964, cg11779113, cg12042264, cg27377213, cg26680502, cg12504877, cg21913888, cg26683005, cg24166457, cg27141915, cg17122157, cg09844573, cg03087897, cg24706505, cg17126555, cg13911392, cg18901104, cg25982880, cg15797834, cg27125093, cg17518965, cg20695297, cg04858553, cg09419005, cg25490145, cg11252953, cg18456621, cg07058988, cg17864646, cg06153925, cg27410601, cg03297901, cg06853339, cg12900649, cg27219182,

cg15759721, cg27023597, cg02782634, cg18942579, cg01409343, cg10530767, cg26112797, cg00253248, cg01722297, cg22589778, cg07137244, cg04147906, cg23878564, cg07860918, cg00206490, cg07644807, cg00558804, cg05304979, cg27598656, cg03549146, cg22190721, cg01660934, cg02358862, cg23093496, cg07641284, cg01681367, cg26769927, cg08480068, cg02914427, cg03653601, cg01990910, cg00933696, cg09866569, cg20357538, cg22460896, cg07116712, cg10186131, cg06380123, cg18610205, cg12353452, cg10590292, cg00037681, cg05596756, cg03569637, cg02522196, cg11655490, cg19693177, cg26363363, cg21249754, cg23147227, cg01657186, cg23764129, cg04514998, cg07332880, cg16061668, cg25574765, cg14088196, cg03758697, cg05398700, cg14058476, cg18158859, cg19300307, cg18842353, cg10732611, cg24480810, cg02053964, cg25922751, cg25954028, cg14642045, cg24165921, cg18215449, cg16402452, cg21376733, cg16509569, cg08075204, cg14556909, cg07119472, cg14999168, cg09399878, cg02874908, cg10542975, cg15698795, cg11791526, cg00862408, cg16260696, cg00220455, cg20826709, cg11436362, cg13924996, cg07420137, cg24301930, cg13395086, cg20136100, cg09153080, cg09902130, cg07380416, cg27284288, cg13912307, cg10511890, cg00242035, cg04314978, cg25225070, cg20411756, cg24247537, cg04330884, cg23130731, cg04888360, cg00907272, cg05979232, cg00025044, cg04441857, cg09684112, cg27388962, cg05931497, cg13408086, cg13555415, cg22552736, cg16191087, cg13925432, cg13464240, cg14633252, cg19252956, cg00015530, cg08632810, cg12737392, cg26769700, cg03218479, cg02609337, cg10351284, cg23554164, cg19021985, cg21031128, cg19421584, cg17984956, cg05177060, cg24107852, cg25652701, cg00282244, cg18887230, cg08486903, cg09335715, cg12629796, cg16454130, cg26433975, cg10673833, cg06787669, cg12192582, cg05098343, cg07573366, cg11105292, cg05287480, cg16748008, cg16644023, cg06488150, cg09450197, cg20336172, cg08858130, cg12098228, cg26811313, cg25432518, cg16622899, cg12359001, cg01209642, cg14564351, cg23429794, cg26401541, cg20046343, cg20847580, cg03431741, cg07417146, cg09001226, cg06482498, cg03891050, cg00899907, cg13597051, cg18113826, cg04859102, cg01620360, cg14083015, cg15046123,

cg03190513, cg01456691, cg17207512, cg20510285, cg01149192, cg05614346, cg06439655, cg11334870, cg08912922, cg23021796, cg24835948, cg10393744, cg07428959, cg17694130, cg03956042, cg19266387, cg13512830, cg19982684, cg22513455, cg07186032, cg08052292, cg27366280, cg06825448, cg25451702, cg08098128, cg13821008, cg27405400, cg09366118, cg15341833, cg02233149, cg14247287, cg23824762, cg01604601, cg05656900, cg08132573, cg24686918, cg05352688, cg18384097, cg16266227, cg19675731, cg21461981, cg25765104, cg26394055, cg20685713, cg23589035, cg01903374, cg23612220, cg26315985, cg18856478, cg23229016, cg21004490, cg24742520, cg23013029, cg19704755, cg07589991, cg10055231, and cg26017930.

20. The computer-implemented method of claim 14, wherein the methylation profile comprises at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, or more biomarkers selected from Table 20.

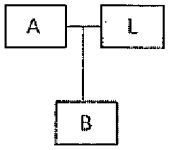
21. The computer-implemented method of claim 14, wherein the cancer type is a solid cancer type or a hematologic malignant cancer type.

22. The computer-implemented method of claim 14, wherein the cancer type comprises acute myeloid leukemia (LAML or AML), acute lymphoblastic leukemia (ALL), adrenocortical carcinoma (ACC), bladder urothelial cancer (BLCA), brain stem glioma, brain lower grade glioma (LGG), brain tumor, breast cancer (BRCA), bronchial tumors, Burkitt lymphoma, cancer of unknown primary site, carcinoid tumor, carcinoma of unknown primary site, central nervous system atypical teratoid/rhabdoid tumor, central nervous system embryonal tumors, cervical squamous cell carcinoma, endocervical adenocarcinoma (CESC) cancer, childhood cancers, cholangiocarcinoma (CHOL), chordoma, chronic lymphocytic leukemia, chronic myelogenous leukemia, chronic myeloproliferative disorders, colon (adenocarcinoma) cancer (COAD), colorectal cancer, craniopharyngioma, cutaneous T-cell lymphoma, endocrine pancreas islet cell tumors, endometrial cancer, ependymoblastoma, ependymoma, esophageal cancer (ESCA), esthesioneuroblastoma, Ewing sarcoma, extracranial germ cell tumor, extragonadal germ cell tumor, extrahepatic bile duct cancer, gallbladder cancer, gastric (stomach) cancer, gastrointestinal carcinoid tumor, gastrointestinal stromal cell tumor, gastrointestinal stromal tumor (GIST), gestational trophoblastic tumor, glioblstoma multiforme glioma GBM), hairy cell leukemia, head and neck cancer (HNSD), heart cancer,

Hodgkin lymphoma, hypopharyngeal cancer, intraocular melanoma, islet cell tumors, Kaposi sarcoma, kidney cancer, Langerhans cell histiocytosis, laryngeal cancer, lip cancer, liver cancer, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma [DLBCL), malignant fibrous histiocytoma bone cancer, medulloblastoma, medullo epithelioma, melanoma, Merkel cell carcinoma, Merkel cell skin carcinoma, mesothelioma (MESO), metastatic squamous neck cancer with occult primary, mouth cancer, multiple endocrine neoplasia syndromes, multiple myeloma, multiple myeloma/plasma cell neoplasm, mycosis fungoides, myelodysplastic syndromes, myeloproliferative neoplasms, nasal cavity cancer, nasopharyngeal cancer, neuroblastoma, Non-Hodgkin lymphoma, nonmelanoma skin cancer, non-small cell lung cancer, oral cancer, oral cavity cancer, oropharyngeal cancer, osteosarcoma, other brain and spinal cord tumors, ovarian cancer, ovarian epithelial cancer, ovarian germ cell tumor, ovarian low malignant potential tumor, pancreatic cancer, papillomatosis, paranasal sinus cancer, parathyroid cancer, pelvic cancer, penile cancer, pharyngeal cancer, pheochromocytoma and paraganglioma (PCPG), pineal parenchymal tumors of intermediate differentiation, pineoblastoma, pituitary tumor, plasma cell neoplasm/multiple myeloma, pleuropulmonary blastoma, primary central nervous system (CNS) lymphoma, primary hepatocellular liver cancer, prostate cancer such as prostate adenocarcinoma (PRAD), rectal cancer, renal cancer, renal cell (kidney) cancer, renal cell cancer, respiratory tract cancer, retinoblastoma, rhabdomyosarcoma, salivary gland cancer, sarcoma (SARC), Sezary syndrome, skin cutaneous melanoma (SKCM), small cell lung cancer, small intestine cancer, soft tissue sarcoma, squamous cell carcinoma, squamous neck cancer, stomach (gastric) cancer, supratentorial primitive neuroectodermal tumors, T-cell lymphoma, testicular cancer testicular germ cell tumors (TGCT), throat cancer, thymic carcinoma, thymoma (THYM), thyroid cancer (THCA), transitional cell cancer, transitional cell cancer of the renal pelvis and ureter, trophoblastic tumor, ureter cancer, urethral cancer, uterine cancer, uterine cancer, uveal melanoma (UVM), vaginal cancer, vulvar cancer, Waldenstrom macroglobulinemia, or Wilm's tumor.

23. The computer-implemented method of claim 14, wherein the biological samples comprise a circulating tumor DNA sample or a tissue sample.

24. A probe panel comprising a plurality of probes, each probe is the probe of Formula I



Formula I

wherein:

A is a first target-binding region;

B is a second target-binding region; and

L is a linker region;

wherein A comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 30 contiguous nucleotides starting at position 1 from the 5' terminus of a sequence selected from SEQ ID NOs: 1-1775; B comprises at least 70%, 80%, 90%, 95%, or 99% sequence identity to at least 12 contiguous nucleotides starting at position 1' from the 3' terminus of the same sequence selected from SEQ ID NOs: 1-1775; L is attached to A; and B is attached to either A or L.

25. The probe panel of claim 24, wherein L is attached to A and B is attached to L.
26. The probe panel of claim 24, wherein the plurality of probes comprises at least 10, 20, 30, 50, 100, or more probes.
27. The probe panel of claim 24, wherein the plurality of probes is used in a solution-based next generation sequencing reaction to generate a CpG methylation data.
28. The probe panel of claim 27, wherein the solution-based next generation sequencing reaction is a droplet digital PCR sequencing method.
29. The probe panel of claim 24, wherein L is between 10 and 60, 15 and 55, 20 and 50, 25 and 45, and 30 and 40 nucleotides in length.
30. The probe panel of claim 24, wherein L further comprises an adaptor region for identification of each probe.