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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 October 2020 (08.10.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/206256 A1

(51) International Patent Classification:
G01N 33/574 (2006.01) *C12N 15/117* (2010.01)
C12N 15/11 (2006.01) *C12Q 1/68* (2018.01)

(21) International Application Number:
PCT/US2020/026581

(22) International Filing Date:
03 April 2020 (03.04.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/828,948 03 April 2019 (03.04.2019) US

(71) Applicants: **MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH** [US/US]; 200 First Street S.W., Rochester, Minnesota 55905 (US). **EXACT SCIENCES DEVELOPMENT COMPANY, LLC** [US/US]; 441 Charmany Drive, Madison, Wisconsin 53719 (US).

(72) Inventors: **TAYLOR, William R.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **KISIEL, John B.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **MAHONEY, Douglas W.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **AHLQUIST, David A.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **ALLAWI, Hatim T.**; c/o Exact Sciences Development Company, LLC, 441 Charmany Drive, Madison, Wisconsin 53719 (US).

(74) Agent: **GOETZ, Robert A.**; Casimir Jones, S.C., 2275 Deming Way, Ste 310, Middleton, Wisconsin 53562 (US).

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

(54) Title: DETECTING PANCREATIC DUCTAL ADENOCARCINOMA IN PLASMA

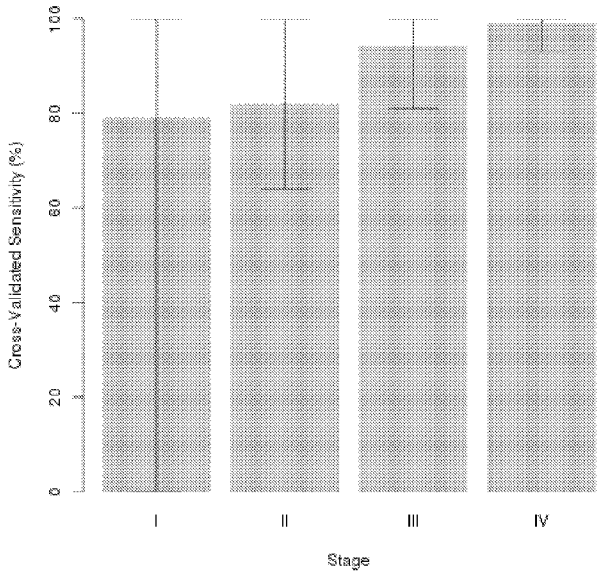


FIG. 2

(57) Abstract: Provided herein is technology for pancreatic ductal adenocarcinoma (PDAC) screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of PDAC.

WO 2020/206256 A1

[Continued on next page]

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

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

DETECTING PANCREATIC DUCTAL ADENOCARCINOMA IN PLASMA

FIELD OF INVENTION

Provided herein is technology for pancreatic ductal adenocarcinoma (PDAC) screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of PDAC.

BACKGROUND

Pancreatic ductal adenocarcinoma (PDAC) is one of the most aggressive solid malignancies. Despite quite a low incidence, it remains the fourth leading cause of cancer-related deaths in the modern world, mainly because of dismal diagnosis (see, Garrido-Laguna I., et al., Nat. Rev. Clin. Oncol. 2015;12:319–334). In the last decades, significant improvements have been achieved in the screening and therapy of different solid cancers, highly incrementing patients' chance for cure. Nevertheless, despite the advancement in pancreatic cancer research, the mortality to incidence ratio has not experienced significant revision over the last few decades. The five-year survival rate remains just around 5–7% and one-year survival is achieved in less than 20% of cases (see, Vincent A., et al., Lancet. 2011;378:607–620). This grim prognosis is mainly caused by the lack of visible and distinctive symptoms and reliable biomarkers for early diagnosis as well as aggressive metastatic spread leading to poor response to treatments (see, Maitra A., Hruban R.H. Annu. Rev. Pathol. 2008;3:157–188).

Improved methods for detecting PDAC and various subtypes of PDAC are needed. The present invention addresses these needs.

SUMMARY

Methylated DNA has been studied as a potential class of biomarkers in the tissues of most tumor types. In many instances, DNA methyltransferases add a methyl group to DNA at cytosine-phosphate-guanine (CpG) island sites as an epigenetic control of gene expression. In a biologically attractive mechanism, acquired methylation events in promoter regions of tumor suppressor genes are thought to silence expression, thus contributing to oncogenesis. DNA methylation may be a more chemically and biologically stable diagnostic tool than RNA or protein expression (Laird (2010) Nat Rev Genet 11: 191–203). Furthermore, in other cancers like sporadic colon cancer, methylation markers offer excellent specificity and are

more broadly informative and sensitive than are individual DNA mutations (Zou et al (2007) Cancer Epidemiol Biomarkers Prev 16: 2686–96).

Analysis of CpG islands has yielded important findings when applied to animal models and human cell lines. For example, Zhang and colleagues found that amplicons from different parts of the same CpG island may have different levels of methylation (Zhang et al. (2009) PLoS Genet 5: e1000438). Further, methylation levels were distributed bi-modally between highly methylated and unmethylated sequences, further supporting the binary switch-like pattern of DNA methyltransferase activity (Zhang et al. (2009) PLoS Genet 5: e1000438). Analysis of murine tissues in vivo and cell lines in vitro demonstrated that only about 0.3% of high CpG density promoters (HCP, defined as having >7% CpG sequence within a 300 base pair region) were methylated, whereas areas of low CpG density (LCP, defined as having <5% CpG sequence within a 300 base pair region) tended to be frequently methylated in a dynamic tissue-specific pattern (Meissner et al. (2008) Nature 454: 766–70). HCPs include promoters for ubiquitous housekeeping genes and highly regulated developmental genes. Among the HCP sites methylated at >50% were several established markers such as Wnt 2, NDRG2, SFRP2, and BMP3 (Meissner et al. (2008) Nature 454: 766–70).

Epigenetic methylation of DNA at cytosine-phosphate-guanine (CpG) island sites by DNA methyltransferases has been studied as a potential class of biomarkers in the tissues of most tumor types. In a biologically attractive mechanism, acquired methylation events in promotor regions of tumor suppressor genes are thought to silence expression, contributing to oncogenesis. DNA methylation may be a more chemically and biologically stable diagnostic tool than RNA or protein expression. Furthermore, in other cancers like sporadic colon cancer, aberrant methylation markers are more broadly informative and sensitive than are individual DNA mutations and offer excellent specificity.

Several methods are available to search for novel methylation markers. While micro-array-based interrogation of CpG methylation is a reasonable, high-throughput approach, this strategy is biased towards known regions of interest, mainly established tumor suppressor promotors. Alternative methods for genome-wide analysis of DNA methylation have been developed in the last decade. There are three basic approaches. The first employs digestion of DNA by restriction enzymes which recognize specific methylated sites, followed by several possible analytic techniques which provide methylation data limited to the enzyme recognition site or the primers used to amplify the DNA in quantification steps (such as

methylation-specific PCR; MSP). A second approach enriches methylated fractions of genomic DNA using anti-bodies directed to methyl-cytosine or other methylation-specific binding domains followed by microarray analysis or sequencing to map the fragment to a reference genome. This approach does not provide single nucleotide resolution of all methylated sites within the fragment. A third approach begins with bisulfite treatment of the DNA to convert all unmethylated cytosines to uracil, followed by restriction enzyme digestion and complete sequencing of all fragments after coupling to an adapter ligand. The choice of restriction enzymes can enrich the fragments for CpG dense regions, reducing the number of redundant sequences which may map to multiple gene positions during analysis.

RRBS yields CpG methylation status data at single nucleotide resolution of 80-90% of all CpG islands and a majority of tumor suppressor promoters at medium to high read coverage. In cancer case - control studies, analysis of these reads results in the identification of differentially methylated regions (DMRs). In previous RRBS analysis of pancreatic cancer specimens, hundreds of DMRs were uncovered, many of which had never been associated with carcinogenesis and many of which were unannotated. Further validation studies on independent tissue samples sets confirmed marker CpGs which were 100% sensitive and specific in terms of performance.

Provided herein is technology for PDAC screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of PDAC.

Indeed, as described in Example I experiments conducted during the course for identifying embodiments for the present invention identified a novel set of differentially methylated regions (DMRs) for discriminating PDAC from non-neoplastic control DNA within tissue and plasma samples.

Such experiments list and describe 13 DNA methylation markers (AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781) distinguishing a) PDAC from non-neoplastic control within plasma samples (see, Table 3, Example I), and b) PDAC tissue from benign pancreatic tissue (see, Table 4, Example 1).

Such experiments identified the following markers and/or panels of markers for detecting PDAC in blood samples (e.g., plasma samples, whole blood samples, leukocyte samples, serum samples):

- AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Table 3, Example 1).

Such experiments identified the following markers and/or panels of markers capable of distinguishing PDAC tissue from benign pancreatic tissue:

- AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Table 4, Example 1).

As described herein, the technology provides a number of methylated DNA markers and subsets thereof (e.g., sets of 2, 3, 4, 5, 6, 7, 8, or 13 markers) with high discrimination for PDAC overall. Experiments applied a selection filter to candidate markers to identify markers that provide a high signal to noise ratio and a low background level to provide high specificity for purposes of PDAC screening or diagnosis.

In some embodiments, the technology is related to assessing the presence of and methylation state of one or more of the markers identified herein in a biological sample (e.g., pancreatic tissue sample, blood sample). These markers comprise one or more differentially methylated regions (DMR) as discussed herein, e.g., as provided in Tables 1. Methylation state is assessed in embodiments of the technology. As such, the technology provided herein is not restricted in the method by which a gene's methylation state is measured. For example, in some embodiments the methylation state is measured by a genome scanning method. For example, one method involves restriction landmark genomic scanning (Kawai et al. (1994) *Mol. Cell. Biol.* 14: 7421–7427) and another example involves methylation-sensitive arbitrarily primed PCR (Gonzalgo et al. (1997) *Cancer Res.* 57: 594–599). In some embodiments, changes in methylation patterns at specific CpG sites are monitored by digestion of genomic DNA with methylation-sensitive restriction enzymes followed by Southern analysis of the regions of interest (digestion-Southern method). In some embodiments, analyzing changes in methylation patterns involves a PCR-based process that involves digestion of genomic DNA with methylation-sensitive restriction enzymes or methylation-dependent restriction enzymes prior to PCR amplification (Singer-Sam et al. (1990) *Nucl. Acids Res.* 18: 687). In addition, other techniques have been reported that utilize bisulfite treatment of DNA as a starting point for methylation analysis. These include methylation-specific PCR (MSP) (Herman et al. (1992) *Proc. Natl. Acad. Sci. USA* 93: 9821–9826) and restriction enzyme digestion of PCR products amplified from bisulfite-converted

DNA (Sadri and Hornsby (1996) *Nucl. Acids Res.* 24: 5058–5059; and Xiong and Laird (1997) *Nucl. Acids Res.* 25: 2532–2534). PCR techniques have been developed for detection of gene mutations (Kuppuswamy et al. (1991) *Proc. Natl. Acad. Sci. USA* 88: 1143–1147) and quantification of allelic-specific expression (Szabo and Mann (1995) *Genes Dev.* 9: 3097–3108; and Singer-Sam et al. (1992) *PCR Methods Appl.* 1: 160–163). Such techniques use internal primers, which anneal to a PCR-generated template and terminate immediately 5' of the single nucleotide to be assayed. Methods using a "quantitative Ms-SNuPE assay" as described in U.S. Pat. No. 7,037,650 are used in some embodiments.

Upon evaluating a methylation state, the methylation state is often expressed as the fraction or percentage of individual strands of DNA that is methylated at a particular site (e.g., at a single nucleotide, at a particular region or locus, at a longer sequence of interest, e.g., up to a ~100-bp, 200-bp, 500-bp, 1000-bp subsequence of a DNA or longer) relative to the total population of DNA in the sample comprising that particular site. Traditionally, the amount of the unmethylated nucleic acid is determined by PCR using calibrators. Then, a known amount of DNA is bisulfite treated and the resulting methylation-specific sequence is determined using either a real-time PCR or other exponential amplification, e.g., a QuARTS assay (e.g., as provided by U.S. Pat. No. 8,361,720; and U.S. Pat. Appl. Pub. Nos. 2012/0122088 and 2012/0122106, incorporated herein by reference).

For example, in some embodiments methods comprise generating a standard curve for the unmethylated target by using external standards. The standard curve is constructed from at least two points and relates the real-time Ct value for unmethylated DNA to known quantitative standards. Then, a second standard curve for the methylated target is constructed from at least two points and external standards. This second standard curve relates the Ct for methylated DNA to known quantitative standards. Next, the test sample Ct values are determined for the methylated and unmethylated populations and the genomic equivalents of DNA are calculated from the standard curves produced by the first two steps. The percentage of methylation at the site of interest is calculated from the amount of methylated DNAs relative to the total amount of DNAs in the population, e.g., (number of methylated DNAs) / (the number of methylated DNAs + number of unmethylated DNAs) × 100.

Also provided herein are compositions and kits for practicing the methods. For example, in some embodiments, reagents (e.g., primers, probes) specific for one or more markers are provided alone or in sets (e.g., sets of primers pairs for amplifying a plurality of markers). Additional reagents for conducting a detection assay may also be provided (e.g.,

enzymes, buffers, positive and negative controls for conducting QuARTS, PCR, sequencing, bisulfite, or other assays). In some embodiments, the kits contain a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent). In some embodiments, the kits containing one or more reagent necessary, sufficient, or useful for conducting a method are provided. Also provided are reactions mixtures containing the reagents. Further provided are master mix reagent sets containing a plurality of reagents that may be added to each other and/or to a test sample to complete a reaction mixture.

In some embodiments, the technology described herein is associated with a programmable machine designed to perform a sequence of arithmetic or logical operations as provided by the methods described herein. For example, some embodiments of the technology are associated with (e.g., implemented in) computer software and/or computer hardware. In one aspect, the technology relates to a computer comprising a form of memory, an element for performing arithmetic and logical operations, and a processing element (e.g., a microprocessor) for executing a series of instructions (e.g., a method as provided herein) to read, manipulate, and store data. In some embodiments, a microprocessor is part of a system for determining a methylation state (e.g., of one or more DMR, e.g., DMR 1-13 as provided in Table 1); comparing methylation states (e.g., of one or more DMR, e.g., DMR 1-13 as provided in Table 1); generating standard curves; determining a Ct value; calculating a fraction, frequency, or percentage of methylation (e.g., of one or more DMR, e.g., DMR 1-13 as provided in Table 1); identifying a CpG island; determining a specificity and/or sensitivity of an assay or marker; calculating an ROC curve and an associated AUC; sequence analysis; all as described herein or is known in the art.

In some embodiments, a microprocessor or computer uses methylation state data in an algorithm to predict a site of a cancer.

In some embodiments, a software or hardware component receives the results of multiple assays and determines a single value result to report to a user that indicates a cancer risk based on the results of the multiple assays (e.g., determining the methylation state of multiple DMR, e.g., as provided in Table 1). Related embodiments calculate a risk factor based on a mathematical combination (e.g., a weighted combination, a linear combination) of the results from multiple assays, e.g., determining the methylation states of multiple markers (such as multiple DMR, e.g., as provided in Table 1). In some embodiments, the methylation state of a DMR defines a dimension and may have values in a multidimensional space and the

coordinate defined by the methylation states of multiple DMR is a result, e.g., to report to a user, e.g., related to a cancer risk.

Some embodiments comprise a storage medium and memory components. Memory components (e.g., volatile and/or nonvolatile memory) find use in storing instructions (e.g., an embodiment of a process as provided herein) and/or data (e.g., a work piece such as methylation measurements, sequences, and statistical descriptions associated therewith). Some embodiments relate to systems also comprising one or more of a CPU, a graphics card, and a user interface (e.g., comprising an output device such as display and an input device such as a keyboard).

Programmable machines associated with the technology comprise conventional extant technologies and technologies in development or yet to be developed (e.g., a quantum computer, a chemical computer, a DNA computer, an optical computer, a spintronics based computer, etc.).

In some embodiments, the technology comprises a wired (e.g., metallic cable, fiber optic) or wireless transmission medium for transmitting data. For example, some embodiments relate to data transmission over a network (e.g., a local area network (LAN), a wide area network (WAN), an ad-hoc network, the internet, etc.). In some embodiments, programmable machines are present on such a network as peers and in some embodiments the programmable machines have a client/server relationship.

In some embodiments, data are stored on a computer-readable storage medium such as a hard disk, flash memory, optical media, a floppy disk, etc.

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

For example, some embodiments provide a computer that includes a computer-readable medium. The embodiment includes a random access memory (RAM) coupled to a processor. The processor executes computer-executable program instructions stored in memory. Such processors may include a microprocessor, an ASIC, a state machine, or other

processor, and can be any of a number of computer processors, such as processors from Intel Corporation of Santa Clara, California and Motorola Corporation of Schaumburg, Illinois. Such processors include, or may be in communication with, media, for example computer-readable media, which stores instructions that, when executed by the processor, cause the processor to perform the steps described herein.

Embodiments of computer-readable media include, but are not limited to, an electronic, optical, magnetic, or other storage or transmission device capable of providing a processor with computer-readable instructions. Other examples of suitable media include, but are not limited to, a floppy disk, CD-ROM, DVD, magnetic disk, memory chip, ROM, RAM, an ASIC, a configured processor, all optical media, all magnetic tape or other magnetic media, or any other medium from which a computer processor can read instructions. Also, various other forms of computer-readable media may transmit or carry instructions to a computer, including a router, private or public network, or other transmission device or channel, both wired and wireless. The instructions may comprise code from any suitable computer-programming language, including, for example, C, C++, C#, Visual Basic, Java, Python, Perl, and JavaScript.

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

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

Accordingly, provided herein is technology related to a method of screening for PDAC in a sample obtained from a subject, the method comprising assaying a methylation state of a marker in a sample obtained from a subject (e.g., pancreatic tissue) (e.g., a blood sample) and identifying the subject as having PDAC when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have PDAC, wherein the marker comprises a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1–13 as provided in Table 1.

In some embodiments wherein the sample obtained from the subject is a blood sample (e.g., plasma sample, whole blood sample, leukocyte sample, serum sample) and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have PDAC indicates the subject has PDAC: AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Table 3, Example 1).

In some embodiments wherein the sample obtained from the subject is pancreatic tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have PDAC indicates the subject has PDAC: AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Table 4, Example 1).

The technology is further related to identifying and discriminating PDAC from blood samples and/or tissue samples. Some embodiments provide methods comprising assaying a plurality of markers (e.g., comprising assaying 2 to 13, 3 to 13, 4 to 13, 5 to 13, 6 to 13, 7 to 13, 8 to 13, 9 to 13, 10 to 13, 11 to 13, 12 to 13) (e.g., comprising assaying no more 13 markers; comprising assaying 13 or fewer markers) (e.g., comprising assaying no more than 12 markers, 11 markers, 10 markers, 9 markers, 8 markers, 7 markers, 6 markers, 5 markers, 4 markers, 3 markers, 2 markers).

The technology is not limited in the methylation state assessed. In some embodiments assessing the methylation state of the marker in the sample comprises determining the methylation state of one base. In some embodiments, assaying the methylation state of the marker in the sample comprises determining the extent of methylation at a plurality of bases. Moreover, in some embodiments the methylation state of the marker comprises an increased methylation of the marker relative to a normal methylation state of the marker. In some embodiments, the methylation state of the marker comprises a decreased methylation of the

marker relative to a normal methylation state of the marker. In some embodiments the methylation state of the marker comprises a different pattern of methylation of the marker relative to a normal methylation state of the marker.

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

The technology is not limited by sample type. For example, in some embodiments the sample is a stool sample, a tissue sample (e.g., pancreatic tissue sample), a blood sample (e.g., plasma, leukocyte, serum, whole blood), an excretion, or a urine sample.

Furthermore, the technology is not limited in the method used to determine methylation state. In some embodiments the assaying comprises using methylation specific polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation specific nuclease, mass-based separation, or target capture. In some embodiments, the assaying comprises use of a methylation specific oligonucleotide. In some embodiments, the technology uses massively parallel sequencing (e.g., next-generation sequencing) to determine methylation state, e.g., sequencing-by-synthesis, real-time (e.g., single-molecule) sequencing, bead emulsion sequencing, nanopore sequencing, etc.

The technology provides reagents for detecting a DMR, e.g., in some embodiments are provided a set of oligonucleotides comprising the sequences provided by SEQ ID NO: 1–13 (see, Table 1). In some embodiments are provided an oligonucleotide comprising a sequence complementary to a chromosomal region having a base in a DMR, e.g., an oligonucleotide sensitive to methylation state of a DMR.

The technology provides various panels of markers use for identifying PDAC, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Tables 3 and/or 4, Example 1).

Kit embodiments are provided, e.g., a kit comprising a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent); and a control nucleic acid comprising a sequence from a DMR selected from a group consisting of DMR 1–13 (from Table 1) and having a methylation state associated with a subject who does not have PDAC. In some embodiments, kits comprise a bisulfite reagent and an oligonucleotide as described

herein. In some embodiments, kits comprise a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent); and a control nucleic acid comprising a sequence from a DMR selected from a group consisting of DMR 1–13 (from Table 1) and having a methylation state associated with a subject who has PDAC. Some kit embodiments comprise a sample collector for obtaining a sample from a subject (e.g., a stool sample; pancreatic tissue sample; blood sample); a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent); and an oligonucleotide as described herein.

The technology is related to embodiments of compositions (e.g., reaction mixtures). In some embodiments are provided a composition comprising a nucleic acid comprising a DMR and a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent). Some embodiments provide a composition comprising a nucleic acid comprising a DMR and an oligonucleotide as described herein. Some embodiments provide a composition comprising a nucleic acid comprising a DMR and a methylation-sensitive restriction enzyme. Some embodiments provide a composition comprising a nucleic acid comprising a DMR and a polymerase.

Additional related method embodiments are provided for screening for PDAC in a sample obtained from a subject (e.g., pancreatic tissue sample; blood sample; stool sample), e.g., a method comprising determining a methylation state of a marker in the sample comprising a base in a DMR that is one or more of DMR 1–13 (from Table 1); comparing the methylation state of the marker from the subject sample to a methylation state of the marker from a normal control sample from a subject who does not have PDAC; and determining a confidence interval and/or a p value of the difference in the methylation state of the subject sample and the normal control sample. In some embodiments, the confidence interval is 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% or 99.99% and the p value is 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, or 0.0001. Some embodiments of methods provide steps of reacting a nucleic acid comprising a DMR with a reagent capable of modifying nucleic acid in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) to produce, for example, nucleic acid modified in a methylation-specific manner; sequencing the nucleic acid modified

in a methylation-specific manner to provide a nucleotide sequence of the nucleic acid modified in a methylation-specific manner; comparing the nucleotide sequence of the nucleic acid modified in a methylation-specific manner with a nucleotide sequence of a nucleic acid comprising the DMR from a subject who does not have PDAC to identify differences in the two sequences; and identifying the subject as having PDAC when a difference is present.

Systems for screening for PDAC in a sample obtained from a subject are provided by the technology. Exemplary embodiments of systems include, e.g., a system for screening for PDAC in a sample obtained from a subject (e.g., pancreatic tissue sample; plasma sample; stool sample), the system comprising an analysis component configured to determine the methylation state of a sample, a software component configured to compare the methylation state of the sample with a control sample or a reference sample methylation state recorded in a database, and an alert component configured to alert a user of a PDAC-associated methylation state. An alert is determined in some embodiments by a software component that receives the results from multiple assays (e.g., determining the methylation states of multiple markers, e.g., DMR, e.g., as provided in Table 1) and calculating a value or result to report based on the multiple results. Some embodiments provide a database of weighted parameters associated with each DMR provided herein for use in calculating a value or result and/or an alert to report to a user (e.g., such as a physician, nurse, clinician, etc.). In some embodiments all results from multiple assays are reported and in some embodiments one or more results are used to provide a score, value, or result based on a composite of one or more results from multiple assays that is indicative of a cancer risk in a subject.

In some embodiments of systems, a sample comprises a nucleic acid comprising a DMR. In some embodiments the system further comprises a component for isolating a nucleic acid, a component for collecting a sample such as a component for collecting a stool sample. In some embodiments, the system comprises nucleic acid sequences comprising a DMR. In some embodiments the database comprises nucleic acid sequences from subjects who do not have PDAC. Also provided are nucleic acids, e.g., a set of nucleic acids, each nucleic acid having a sequence comprising a DMR. In some embodiments the set of nucleic acids wherein each nucleic acid has a sequence from a subject who does not have PDAC. Related system embodiments comprise a set of nucleic acids as described and a database of nucleic acid sequences associated with the set of nucleic acids. Some embodiments further comprise a reagent capable of modifying DNA in a methylation-specific manner (e.g., a

methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent). And, some embodiments further comprise a nucleic acid sequencer.

In certain embodiments, methods for characterizing a sample (e.g., pancreatic tissue sample; blood sample; stool sample) from a human patient are provided. For example, in some embodiments such embodiments comprise obtaining DNA from a sample of a human patient; assaying a methylation state of a DNA methylation marker comprising a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1–13 from Table 1; and comparing the assayed methylation state of the one or more DNA methylation markers with methylation level references for the one or more DNA methylation markers for human patients not having PDAC.

Such methods are not limited to a particular type of sample from a human patient. In some embodiments, the sample is a pancreatic tissue sample. In some embodiments, the sample is a plasma sample. In some embodiments, the sample is a stool sample, a tissue sample, a pancreatic tissue sample, a blood sample (e.g., leukocyte sample, plasma sample, whole blood sample, serum sample), or a urine sample.

In some embodiments, such methods comprise assaying a plurality of DNA methylation markers (e.g., comprising assaying 2 to 13, 3 to 13, 4 to 13, 5 to 13, 6 to 13, 7 to 13, 8 to 13, 9 to 13, 10 to 13, 11 to 13, 12 to 13) (e.g., comprising assaying no more 13 markers; comprising assaying 13 or fewer markers) (e.g., comprising assaying no more than 12 markers, 11 markers, 10 markers, 9 markers, 8 markers, 7 markers, 6 markers, 5 markers, 4 markers, 3 markers, 2 markers). In some embodiments, such methods comprise assaying the methylation state of the one or more DNA methylation markers in the sample comprises determining the methylation state of one base. In some embodiments, such methods comprise assaying the methylation state of the one or more DNA methylation markers in the sample comprises determining the extent of methylation at a plurality of bases. In some embodiments, such methods comprise assaying a methylation state of a forward strand or assaying a methylation state of a reverse strand.

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

In some embodiments, the assaying comprises using methylation specific polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation specific nuclease, mass-based separation, or target capture.

In some embodiments, the assaying comprises use of a methylation specific oligonucleotide. In some embodiments, the methylation specific oligonucleotide is selected from the group consisting of SEQ ID NO: 1–13 (Table 1).

In some embodiments, a chromosomal region having an annotation selected from the group consisting of AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Table 1, Example 1) comprises the DNA methylation marker.

In some embodiments, such methods comprise determining the methylation state of two DNA methylation markers. In some embodiments, such methods comprise determining the methylation state of a pair of DNA methylation markers provided in a row of Table 1.

In certain embodiments, the technology provides methods for characterizing a sample (e.g., pancreatic tissue sample; leukocyte sample; plasma sample; whole blood sample; serum sample; stool sample) obtained from a human patient. In some embodiments, such methods comprise determining a methylation state of a DNA methylation marker in the sample comprising a base in a DMR selected from a group consisting of DMR 1–13 from Table 1; comparing the methylation state of the DNA methylation marker from the patient sample to a methylation state of the DNA methylation marker from a normal control sample from a human subject who does not have PDAC; and determining a confidence interval and/or a *p* value of the difference in the methylation state of the human patient and the normal control sample. In some embodiments, the confidence interval is 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% or 99.99% and the *p* value is 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, or 0.0001.

In certain embodiments, the technology provides methods for characterizing a sample obtained from a human subject (e.g., pancreatic tissue sample; leukocyte sample; plasma sample; whole blood sample; serum sample; stool sample), the method comprising reacting a nucleic acid comprising a DMR with a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) to produce nucleic acid modified in a methylation-specific manner; sequencing the nucleic acid modified in a methylation-specific manner to provide a nucleotide sequence of the nucleic acid modified in a methylation-specific manner;

comparing the nucleotide sequence of the nucleic acid modified in a methylation-specific manner with a nucleotide sequence of a nucleic acid comprising the DMR from a subject who does not have PDAC to identify differences in the two sequences.

In certain embodiments, the technology provides systems for characterizing a sample obtained from a human subject (e.g., pancreatic tissue sample; plasma sample; stool sample), the system comprising an analysis component configured to determine the methylation state of a sample, a software component configured to compare the methylation state of the sample with a control sample or a reference sample methylation state recorded in a database, and an alert component configured to determine a single value based on a combination of methylation states and alert a user of a PDAC-associated methylation state. In some embodiments, the sample comprises a nucleic acid comprising a DMR.

In some embodiments, such systems further comprise a component for isolating a nucleic acid. In some embodiments, such systems further comprise a component for collecting a sample.

In some embodiments, the sample is a stool sample, a tissue sample, a pancreatic tissue sample, a blood sample (e.g., plasma sample, leukocyte sample, whole blood sample, serum sample), or a urine sample.

In some embodiments, the database comprises nucleic acid sequences comprising a DMR. In some embodiments, the database comprises nucleic acid sequences from subjects who do not have PDAC.

Additional embodiments will be apparent to persons skilled in the relevant art based on the teachings contained herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1: Marker chromosomal regions used for the 13 methylated DNA markers recited in Table 1 and related primer and probe information.

Fig. 2: Cross-validated sensitivity of a methylated DNA marker-CA 19-9 panel across PDAC stages at 92% specificity

Fig. 3: Cross-validated ROC curve for methylated DNA marker panel alone, CA 19-9 alone, combined panel for discrimination of PDAC.

DEFINITIONS

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

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

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

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

As used herein, a "nucleic acid" or "nucleic acid molecule" generally refers to any ribonucleic acid or deoxyribonucleic acid, which may be unmodified or modified DNA or RNA. "Nucleic acids" include, without limitation, single- and double-stranded nucleic acids. As used herein, the term "nucleic acid" also includes DNA as described above that contains one or more modified bases. Thus, DNA with a backbone modified for stability or for other reasons is a "nucleic acid". The term "nucleic acid" as it is used herein embraces such chemically, enzymatically, or metabolically modified forms of nucleic acids, as well as the chemical forms of DNA characteristic of viruses and cells, including for example, simple and complex cells.

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

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

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

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

The term “gene” also encompasses the coding regions of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends, e.g., for a distance of about 1 kb on either end, such that the gene corresponds to the length of the full-length mRNA (e.g., comprising coding, regulatory, structural and other sequences). The sequences that are located 5' of the coding region and that are present on the mRNA are

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

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

The term "wild-type" when made in reference to a gene refers to a gene that has the characteristics of a gene isolated from a naturally occurring source. The term "wild-type" when made in reference to a gene product refers to a gene product that has the characteristics of a gene product isolated from a naturally occurring source. The term "naturally-occurring" as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by the hand of a person in the laboratory is naturally-occurring. A wild-type gene is often that gene or allele that is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" when made in reference to a gene or to a gene product refers, respectively, to a gene or to a gene product that displays modifications in sequence and/or functional properties (e.g., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

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

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

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (e.g., replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (e.g., synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

The term "amplifying" or "amplification" in the context of nucleic acids refers to the production of multiple copies of a polynucleotide, or a portion of the polynucleotide, typically starting from a small amount of the polynucleotide (*e.g.*, a single polynucleotide molecule), where the amplification products or amplicons are generally detectable. Amplification of polynucleotides encompasses a variety of chemical and enzymatic processes. The generation of multiple DNA copies from one or a few copies of a target or template DNA molecule during a polymerase chain reaction (PCR) or a ligase chain reaction (LCR; see, *e.g.*, U.S. Patent No. 5,494,810; herein incorporated by reference in its entirety) are forms of amplification. Additional types of amplification include, but are not limited to, allele-specific PCR (see, *e.g.*, U.S. Patent No. 5,639,611; herein incorporated by reference in its entirety), assembly PCR (see, *e.g.*, U.S. Patent No. 5,965,408; herein incorporated by reference in its entirety), helicase-dependent amplification (see, *e.g.*, U.S. Patent No. 7,662,594; herein incorporated by reference in its entirety), hot-start PCR (see, *e.g.*, U.S. Patent Nos. 5,773,258 and 5,338,671; each herein incorporated by reference in their entireties), intersequence-specific PCR, inverse PCR (see, *e.g.*, Triglia, *et al.*(1988) Nucleic Acids Res., 16:8186; herein incorporated by reference in its entirety), ligation-mediated PCR (see, *e.g.*, Guilfoyle, R. *et al.*, Nucleic Acids Research, 25:1854-1858 (1997); U.S. Patent No.

5,508,169; each of which are herein incorporated by reference in their entireties), methylation-specific PCR (see, *e.g.*, Herman, *et al.*, (1996) PNAS 93(13) 9821-9826; herein incorporated by reference in its entirety), miniprimer PCR, multiplex ligation-dependent probe amplification (see, *e.g.*, Schouten, *et al.*, (2002) Nucleic Acids Research 30(12): e57; herein incorporated by reference in its entirety), multiplex PCR (see, *e.g.*, Chamberlain, *et al.*, (1988) Nucleic Acids Research 16(23) 11141-11156; Ballabio, *et al.*, (1990) Human Genetics 84(6) 571-573; Hayden, *et al.*, (2008) BMC Genetics 9:80; each of which are herein incorporated by reference in their entireties), nested PCR, overlap-extension PCR (see, *e.g.*, Higuchi, *et al.*, (1988) Nucleic Acids Research 16(15) 7351-7367; herein incorporated by reference in its entirety), real time PCR (see, *e.g.*, Higuchi, *et al.*, (1992) Biotechnology 10:413-417; Higuchi, *et al.*, (1993) Biotechnology 11:1026-1030; each of which are herein incorporated by reference in their entireties), reverse transcription PCR (see, *e.g.*, Bustin, S.A. (2000) J. Molecular Endocrinology 25:169-193; herein incorporated by reference in its entirety), solid phase PCR, thermal asymmetric interlaced PCR, and Touchdown PCR (see, *e.g.*, Don, *et al.*, Nucleic Acids Research (1991) 19(14) 4008; Roux, K. (1994) Biotechniques 16(5) 812-814; Hecker, *et al.*, (1996) Biotechniques 20(3) 478-485; each of which are herein incorporated by reference in their entireties). Polynucleotide amplification also can be accomplished using digital PCR (see, *e.g.*, Kalinina, *et al.*, Nucleic Acids Research. 25; 1999-2004, (1997); Vogelstein and Kinzler, Proc Natl Acad Sci USA. 96; 9236-41, (1999); International Patent Publication No. WO05023091A2; US Patent Application Publication No. 20070202525; each of which are incorporated herein by reference in their entireties).

The term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic or other DNA or RNA, without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension

constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" ("PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified" and are "PCR products" or "amplicons." Those of skill in the art will understand the term "PCR" encompasses many variants of the originally described method using, *e.g.*, real time PCR, nested PCR, reverse transcription PCR (RT-PCR), single primer and arbitrarily primed PCR, *etc.*

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q-beta replicase, MDV-1 RNA is the specific template for the replicase (Kacian et al., Proc. Natl. Acad. Sci. USA, 69:3038 [1972]). Other nucleic acid will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin et al, Nature, 228:227 [1970]). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace (1989) Genomics 4:560). Finally, thermostable template-dependant DNA polymerases (e.g., Taq and Pfu DNA polymerases), by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H. A. Erlich (ed.), PCR Technology, Stockton Press [1989]).

As used herein, the term "nucleic acid detection assay" refers to any method of determining the nucleotide composition of a nucleic acid of interest. Nucleic acid detection assay include but are not limited to, DNA sequencing methods, probe hybridization methods, structure specific cleavage assays (*e.g.*, the INVADER assay, (Hologic, Inc.) and are described, *e.g.*, in U.S. Patent Nos. 5,846,717, 5,985,557, 5,994,069, 6,001,567, 6,090,543, and 6,872,816; Lyamichev et al., Nat. Biotech., 17:292 (1999), Hall et al., PNAS, USA,

97:8272 (2000), and US Pat. No. 9,096,893, each of which is herein incorporated by reference in its entirety for all purposes); enzyme mismatch cleavage methods (*e.g.*, Variagenics, U.S. Pat. Nos. 6,110,684, 5,958,692, 5,851,770, herein incorporated by reference in their entireties); polymerase chain reaction (PCR), described above; branched hybridization methods (*e.g.*, Chiron, U.S. Pat. Nos. 5,849,481, 5,710,264, 5,124,246, and 5,624,802, herein incorporated by reference in their entireties); rolling circle replication (*e.g.*, U.S. Pat. Nos. 6,210,884, 6,183,960 and 6,235,502, herein incorporated by reference in their entireties); NASBA (*e.g.*, U.S. Pat. No. 5,409,818, herein incorporated by reference in its entirety); molecular beacon technology (*e.g.*, U.S. Pat. No. 6,150,097, herein incorporated by reference in its entirety); E-sensor technology (Motorola, U.S. Pat. Nos. 6,248,229, 6,221,583, 6,013,170, and 6,063,573, herein incorporated by reference in their entireties); cycling probe technology (*e.g.*, U.S. Pat. Nos. 5,403,711, 5,011,769, and 5,660,988, herein incorporated by reference in their entireties); Dade Behring signal amplification methods (*e.g.*, U.S. Pat. Nos. 6,121,001, 6,110,677, 5,914,230, 5,882,867, and 5,792,614, herein incorporated by reference in their entireties); ligase chain reaction (*e.g.*, Baranay Proc. Natl. Acad. Sci USA 88, 189-93 (1991)); and sandwich hybridization methods (*e.g.*, U.S. Pat. No. 5,288,609, herein incorporated by reference in its entirety).

The term "amplifiable nucleic acid" refers to a nucleic acid that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

The term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

The term "primer" refers to an oligonucleotide, whether occurring naturally as, *e.g.*, a nucleic acid fragment from a restriction digest, or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid template strand is induced, (*e.g.*, in the presence of nucleotides and an inducing agent such as a DNA polymerase, and at a suitable temperature and pH). The primer is preferably single stranded

for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer, and the use of the method.

The term "probe" refers to an oligonucleotide (*e.g.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly, or by PCR amplification, that is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification, and isolation of particular gene sequences (*e.g.*, a "capture probe"). It is contemplated that any probe used in the present invention may, in some embodiments, be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

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

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

Accordingly, as used herein a "methylated nucleotide" or a "methylated nucleotide base" refers to the presence of a methyl moiety on a nucleotide base, where the methyl

moiety is not present in a recognized typical nucleotide base. For example, cytosine does not contain a methyl moiety on its pyrimidine ring, but 5-methylcytosine contains a methyl moiety at position 5 of its pyrimidine ring. Therefore, cytosine is not a methylated nucleotide and 5-methylcytosine is a methylated nucleotide. In another example, thymine contains a methyl moiety at position 5 of its pyrimidine ring; however, for purposes herein, thymine is not considered a methylated nucleotide when present in DNA since thymine is a typical nucleotide base of DNA.

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

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

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

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

The methylation status can optionally be represented or indicated by a "methylation value" (e.g., representing a methylation frequency, fraction, ratio, percent, etc.) A methylation value can be generated, for example, by quantifying the amount of intact nucleic acid present following restriction digestion with a methylation dependent restriction enzyme or by comparing amplification profiles after bisulfite reaction or by comparing sequences of

bisulfite-treated and untreated nucleic acids. Accordingly, a value, e.g., a methylation value, represents the methylation status and can thus be used as a quantitative indicator of methylation status across multiple copies of a locus. This is of particular use when it is desirable to compare the methylation status of a sequence in a sample to a threshold or reference value.

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

As such, the methylation state describes the state of methylation of a nucleic acid (e.g., a genomic sequence). In addition, the methylation state refers to the characteristics of a nucleic acid segment at a particular genomic locus relevant to methylation. Such characteristics include, but are not limited to, whether any of the cytosine (C) residues within this DNA sequence are methylated, the location of methylated C residue(s), the frequency or percentage of methylated C throughout any particular region of a nucleic acid, and allelic differences in methylation due to, e.g., difference in the origin of the alleles. The terms "methylation state", "methylation profile", and "methylation status" also refer to the relative concentration, absolute concentration, or pattern of methylated C or unmethylated C throughout any particular region of a nucleic acid in a biological sample. For example, if the cytosine (C) residue(s) within a nucleic acid sequence are methylated it may be referred to as "hypermethylated" or having "increased methylation", whereas if the cytosine (C) residue(s) within a DNA sequence are not methylated it may be referred to as "hypomethylated" or having "decreased methylation". Likewise, if the cytosine (C) residue(s) within a nucleic acid sequence are methylated as compared to another nucleic acid sequence (e.g., from a different region or from a different individual, etc.) that sequence is considered hypermethylated or having increased methylation compared to the other nucleic acid sequence. Alternatively, if the cytosine (C) residue(s) within a DNA sequence are not methylated as compared to another nucleic acid sequence (e.g., from a different region or from a different individual, etc.) that sequence is considered hypomethylated or having decreased methylation compared to the other nucleic acid sequence. Additionally, the term "methylation pattern" as used herein refers to the collective sites of methylated and unmethylated nucleotides over a region of a nucleic acid. Two nucleic acids may have the same or similar methylation frequency or methylation percent but have different methylation patterns when the number of methylated and unmethylated nucleotides are the same or similar throughout the region but the locations

of methylated and unmethylated nucleotides are different. Sequences are said to be "differentially methylated" or as having a "difference in methylation" or having a "different methylation state" when they differ in the extent (e.g., one has increased or decreased methylation relative to the other), frequency, or pattern of methylation. The term "differential methylation" refers to a difference in the level or pattern of nucleic acid methylation in a cancer positive sample as compared with the level or pattern of nucleic acid methylation in a cancer negative sample. It may also refer to the difference in levels or patterns between patients that have recurrence of cancer after surgery versus patients who not have recurrence. Differential methylation and specific levels or patterns of DNA methylation are prognostic and predictive biomarkers, e.g., once the correct cut-off or predictive characteristics have been defined.

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

As used herein a "nucleotide locus" refers to the location of a nucleotide in a nucleic acid molecule. A nucleotide locus of a methylated nucleotide refers to the location of a methylated nucleotide in a nucleic acid molecule.

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

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

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

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

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

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

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

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

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

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

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

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

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

As used herein, a "selected nucleotide" refers to one nucleotide of the four typically occurring nucleotides in a nucleic acid molecule (C, G, T, and A for DNA and C, G, U, and A for RNA), and can include methylated derivatives of the typically occurring nucleotides

(e.g., when C is the selected nucleotide, both methylated and unmethylated C are included within the meaning of a selected nucleotide), whereas a methylated selected nucleotide refers specifically to a methylated typically occurring nucleotide and an unmethylated selected nucleotides refers specifically to an unmethylated typically occurring nucleotide.

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

As used herein, a "different nucleotide" refers to a nucleotide that is chemically different from a selected nucleotide, typically such that the different nucleotide has Watson-Crick base-pairing properties that differ from the selected nucleotide, whereby the typically occurring nucleotide that is complementary to the selected nucleotide is not the same as the typically occurring nucleotide that is complementary to the different nucleotide. For example, when C is the selected nucleotide, U or T can be the different nucleotide, which is exemplified by the complementarity of C to G and the complementarity of U or T to A. As used herein, a nucleotide that is complementary to the selected nucleotide or that is complementary to the different nucleotide refers to a nucleotide that base-pairs, under high stringency conditions, with the selected nucleotide or different nucleotide with higher affinity than the complementary nucleotide's base-paring with three of the four typically occurring nucleotides. An example of complementarity is Watson-Crick base pairing in DNA (e.g., A-T and C-G) and RNA (e.g., A-U and C-G). Thus, for example, G base-pairs, under high stringency conditions, with higher affinity to C than G base-pairs to G, A, or T and, therefore, when C is the selected nucleotide, G is a nucleotide complementary to the selected nucleotide.

As used herein, the "sensitivity" of a given marker (or set of markers used together) refers to the percentage of samples that report a DNA methylation value above a threshold value that distinguishes between neoplastic and non-neoplastic samples. In some embodiments, a positive is defined as a histology-confirmed neoplasia that reports a DNA methylation value above a threshold value (*e.g.*, the range associated with disease), and a false negative is defined as a histology-confirmed neoplasia that reports a DNA methylation value below the threshold value (*e.g.*, the range associated with no disease). The value of sensitivity, therefore, reflects the probability that a DNA methylation measurement for a given marker obtained from a known diseased sample will be in the range of disease-associated measurements. As defined here, the clinical relevance of the calculated sensitivity value represents an estimation of the probability that a given marker would detect the presence of a clinical condition when applied to a subject with that condition.

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

The term "AUC" as used herein is an abbreviation for the "area under a curve". In particular it refers to the area under a Receiver Operating Characteristic (ROC) curve. The ROC curve is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test. It shows the trade-off between sensitivity and specificity depending on the selected cut point (any increase in sensitivity will be accompanied by a decrease in specificity). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test (the larger the area the better; the optimum is 1; a random

test would have a ROC curve lying on the diagonal with an area of 0.5; for reference: J. P. Egan. (1975) *Signal Detection Theory and ROC Analysis*, Academic Press, New York).

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

The term "neoplasm-specific marker," as used herein, refers to any biological material or element that can be used to indicate the presence of a neoplasm. Examples of biological materials include, without limitation, nucleic acids, polypeptides, carbohydrates, fatty acids, cellular components (*e.g.*, cell membranes and mitochondria), and whole cells. In some instances, markers are particular nucleic acid regions (*e.g.*, genes, intragenic regions, specific loci, etc.). Regions of nucleic acid that are markers may be referred to, *e.g.*, as "marker genes," "marker regions," "marker sequences," "marker loci," etc.

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

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

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

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

The term “isolated” when used in relation to a nucleic acid, as in “an isolated oligonucleotide” refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids, such as DNA and RNA, are found in the state they exist in nature. Examples of non-isolated nucleic acids include: a given DNA

sequence (e.g., a gene) found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, isolated nucleic acid encoding a particular protein includes, by way of example, such nucleic acid in cells ordinarily expressing the protein, where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or oligonucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide may be single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide may be double-stranded). An isolated nucleic acid may, after isolation from its natural or typical environment, by be combined with other nucleic acids or molecules. For example, an isolated nucleic acid may be present in a host cell in which into which it has been placed, e.g., for heterologous expression.

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

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

The term "sample" is used in its broadest sense. In one sense it can refer to an animal cell or tissue. In another sense, it refers to a specimen or culture obtained from any source, as

well as biological and environmental samples. Biological samples may be obtained from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. Environmental samples include environmental material such as surface matter, soil, water, and industrial samples. These examples are not to be construed as limiting the sample types applicable to the present invention.

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

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

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

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

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

DETAILED DESCRIPTION

In this detailed description of the various embodiments, for purposes of explanation, numerous specific details are set forth to provide a thorough understanding of the embodiments disclosed. One skilled in the art will appreciate, however, that these various embodiments may be practiced with or without these specific details. In other instances,

structures and devices are shown in block diagram form. Furthermore, one skilled in the art can readily appreciate that the specific sequences in which methods are presented and performed are illustrative and it is contemplated that the sequences can be varied and still remain within the spirit and scope of the various embodiments disclosed herein.

Provided herein is technology for PDAC screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of PDAC. As the technology is described herein, the section headings used are for organizational purposes only and are not to be construed as limiting the subject matter in any way.

Indeed, as described in Example 1, experiments conducted during the course for identifying embodiments for the present invention identified 13 differentially methylated regions (DMRs) for discriminating PDAC from non-neoplastic control DNA.

Such experiments list and describe 13 DNA methylation markers (AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781) distinguishing a) PDAC from non-neoplastic control within plasma samples (see, Table 3, Example I), and b) PDAC tissue from benign pancreatic tissue (see, Table 4, Example 1).

Such experiments identified the following markers and/or panels of markers for detecting PDAC in blood samples (e.g., plasma samples, whole blood samples, leukocyte samples, serum samples):

- AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Table 3, Example 1).

Such experiments identified the following markers and/or panels of markers capable of distinguishing PDAC tissue from benign pancreatic tissue:

AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 (see, Table 4, Example 1).

Although the disclosure herein refers to certain illustrated embodiments, it is to be understood that these embodiments are presented by way of example and not by way of limitation.

In particular aspects, the present technology provides compositions and methods for identifying, determining, and/or classifying a cancer such as PDAC. The methods comprise determining the methylation status of at least one methylation marker in a biological sample isolated from a subject (e.g., stool sample, pancreatic tissue sample, plasma sample), wherein

a change in the methylation state of the marker is indicative of the presence, class, or site of PDAC. Particular embodiments relate to markers comprising a differentially methylated region (DMR, e.g., DMR 1–13, see Table 1) that are used for diagnosis (e.g., screening) of PDAC.

In addition to embodiments wherein the methylation analysis of at least one marker, a region of a marker, or a base of a marker comprising a DMR (e.g., DMR, e.g., DMR 1–13) provided herein and listed in Table 1 is analyzed, the technology also provides panels of markers comprising at least one marker, region of a marker, or base of a marker comprising a DMR with utility for the detection of cancers, in particular PDAC.

Some embodiments of the technology are based upon the analysis of the CpG methylation status of at least one marker, region of a marker, or base of a marker comprising a DMR.

In some embodiments, the present technology provides for the use of a reagent that modifies DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) in combination with one or more methylation assays to determine the methylation status of CpG dinucleotide sequences within at least one marker comprising a DMR (e.g., DMR 1–13, see Table 1). Genomic CpG dinucleotides can be methylated or unmethylated (alternatively known as up- and down-methylated respectively). However, the methods of the present invention are suitable for the analysis of biological samples of a heterogeneous nature, e.g., a low concentration of tumor cells, or biological materials therefrom, within a background of a remote sample (e.g., blood, organ effluent, or stool). Accordingly, when analyzing the methylation status of a CpG position within such a sample one may use a quantitative assay for determining the level (e.g., percent, fraction, ratio, proportion, or degree) of methylation at a particular CpG position.

According to the present technology, determination of the methylation status of CpG dinucleotide sequences in markers comprising a DMR has utility both in the diagnosis and characterization of cancers such as PDAC.

Combinations of markers

In some embodiments, the technology relates to assessing the methylation state of combinations of markers comprising a DMR from Table 1 (e.g., DMR Nos. 1-13). In some embodiments, assessing the methylation state of more than one marker increases the

specificity and/or sensitivity of a screen or diagnostic for identifying a neoplasm in a subject (e.g., PDAC).

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

Methods for assaying methylation state

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

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

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

The bisulfite technique typically involves amplifying short, specific fragments of a known nucleic acid subsequent to a bisulfite treatment, then either assaying the product by sequencing (Olek & Walter (1997) *Nat. Genet.* 17: 275–6) or a primer extension reaction (Gonzalgo & Jones (1997) *Nucleic Acids Res.* 25: 2529–31; WO 95/00669; U.S. Pat. No. 6,251,594) to analyze individual cytosine positions. Some methods use enzymatic digestion (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532–4). Detection by hybridization has also been described in the art (Olek et al., WO 99/28498). Additionally, use of the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg & Clark (1994) *Bioessays* 16: 431–6; Zeschnigk et al. (1997) *Hum Mol Genet.* 6: 387–95; Feil et al. (1994) *Nucleic Acids Res.* 22: 695; Martin et al. (1995) *Gene* 157: 261–4; WO 9746705; WO 9515373).

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

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

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

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

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

covered by, the amplification primers enable methylation-specific selective amplification of a nucleic acid sample.

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

Typical reagents (*e.g.*, as might be found in a typical MethyLight™-based kit) for HeavyMethyl™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, or bisulfite treated DNA sequence or CpG island, *etc.*); blocking oligonucleotides; optimized PCR buffers and deoxynucleotides; and Taq polymerase. MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; U.S. Pat. No. 5,786,146). Briefly, DNA is modified by sodium bisulfite, which converts unmethylated, but not methylated cytosines, to uracil, and the products are subsequently amplified with primers specific for methylated versus unmethylated DNA. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); optimized PCR buffers and deoxynucleotides, and specific probes.

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

both at the level of the amplification process and at the level of the fluorescence detection process.

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

The MethyLight™ process is used with any suitable probe (*e.g.* a "TaqMan®" probe, a Lightcycler® probe, *etc.*) For example, in some applications double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes, *e.g.*, with MSP primers and/or HeavyMethyl blocker oligonucleotides and a TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules and is designed to be specific for a relatively high GC content region so that it melts at about a 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5′ to 3′ endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

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

The QM™ (quantitative methylation) assay is an alternative quantitative test for methylation patterns in genomic DNA samples, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for

unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe, overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing the biased PCR pool with either control oligonucleotides that do not cover known methylation sites (a fluorescence-based version of the HeavyMethyl™ and MSP techniques) or with oligonucleotides covering potential methylation sites.

The QM™ process can be used with any suitable probe, *e.g.*, "TaqMan®" probes, Lightcycler® probes, in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to unbiased primers and the TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about a 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5′ to 3′ endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system. Typical reagents (*e.g.*, as might be found in a typical QM™-based kit) for QM™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); TaqMan® or Lightcycler® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

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

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

Reduced Representation Bisulfite Sequencing (RRBS) begins with bisulfite treatment of nucleic acid to convert all unmethylated cytosines to uracil, followed by restriction enzyme digestion (*e.g.*, by an enzyme that recognizes a site including a CG sequence such as MspI) and complete sequencing of fragments after coupling to an adapter ligand. The choice of restriction enzyme enriches the fragments for CpG dense regions, reducing the number of redundant sequences that may map to multiple gene positions during analysis. As such, RRBS reduces the complexity of the nucleic acid sample by selecting a subset (*e.g.*, by size selection using preparative gel electrophoresis) of restriction fragments for sequencing. As opposed to whole-genome bisulfite sequencing, every fragment produced by the restriction enzyme digestion contains DNA methylation information for at least one CpG dinucleotide. As such, RRBS enriches the sample for promoters, CpG islands, and other genomic features with a high frequency of restriction enzyme cut sites in these regions and thus provides an assay to assess the methylation state of one or more genomic loci.

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

In some embodiments, a quantitative allele-specific real-time target and signal amplification (QuARTS) assay is used to evaluate methylation state. Three reactions sequentially occur in each QuARTS assay, including amplification (reaction 1) and target probe cleavage (reaction 2) in the primary reaction; and FRET cleavage and fluorescent signal generation (reaction 3) in the secondary reaction. When target nucleic acid is amplified

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

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

In some embodiments, fragments of the treated DNA are amplified using sets of primer oligonucleotides according to the present invention (e.g., see Tables 10, 19 and 20) and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). Amplicons are typically 100 to 2000 base pairs in length.

In another embodiment of the method, the methylation status of CpG positions within or near a marker comprising a DMR (e.g., DMR 1–13, Table 1) may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in U.S. Pat. No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primer pairs contain at least one primer that hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the position of the C position in the CpG.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. In some embodiments, the labels are fluorescent labels, radionuclides, or detachable molecule fragments having a typical mass that can be detected in a mass spectrometer. Where said labels are mass labels, some embodiments provide that the labeled amplicons have a single positive or negative net charge, allowing for better delectability in the mass spectrometer. The detection may be carried out and visualized by means of, e.g., matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

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

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

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

In some embodiments, the sample comprises blood, serum, leukocytes, plasma, or saliva. In some embodiments, the subject is human. Such samples can be obtained by any number of means known in the art, such as will be apparent to the skilled person. Cell free or substantially cell free samples can be obtained by subjecting the sample to various techniques known to those of skill in the art which include, but are not limited to, centrifugation and filtration. Although it is generally preferred that no invasive techniques are used to obtain the sample, it still may be preferable to obtain samples such as tissue homogenates, tissue sections, and biopsy specimens. The technology is not limited in the methods used to prepare the samples and provide a nucleic acid for testing. For example, in some embodiments, a DNA is isolated from a stool sample or from blood or from a plasma sample using direct gene capture, *e.g.*, as detailed in U.S. Pat. Nos. 8,808,990 and 9,169,511, and in WO 2012/155072, or by a related method.

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

numbers of test samples. Alternatively, single sample formats could be developed to facilitate immediate treatment and diagnosis in a timely fashion, for example, in ambulatory transport or emergency room settings.

It is contemplated that embodiments of the technology are provided in the form of a kit. The kits comprise embodiments of the compositions, devices, apparatuses, *etc.* described herein, and instructions for use of the kit. Such instructions describe appropriate methods for preparing an analyte from a sample, *e.g.*, for collecting a sample and preparing a nucleic acid from the sample. Individual components of the kit are packaged in appropriate containers and packaging (*e.g.*, vials, boxes, blister packs, ampules, jars, bottles, tubes, and the like) and the components are packaged together in an appropriate container (*e.g.*, a box or boxes) for convenient storage, shipping, and/or use by the user of the kit. It is understood that liquid components (*e.g.*, a buffer) may be provided in a lyophilized form to be reconstituted by the user. Kits may include a control or reference for assessing, validating, and/or assuring the performance of the kit. For example, a kit for assaying the amount of a nucleic acid present in a sample may include a control comprising a known concentration of the same or another nucleic acid for comparison and, in some embodiments, a detection reagent (*e.g.*, a primer) specific for the control nucleic acid. The kits are appropriate for use in a clinical setting and, in some embodiments, for use in a user's home. The components of a kit, in some embodiments, provide the functionalities of a system for preparing a nucleic acid solution from a sample. In some embodiments, certain components of the system are provided by the user.

Methods

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

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from a blood sample (e.g., plasma sample, whole blood sample, leukocyte sample, serum sample) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781, and

2) detecting PDAC (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

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

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from pancreatic tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781, and
2) detecting PDAC (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

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

1) measuring a methylation level for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with a reagent that modifies DNA in a methylation-specific manner (e.g., wherein the reagent is a bisulfite reagent, a methylation-sensitive restriction enzyme, or a methylation-dependent restriction enzyme), wherein the one or more genes is selected from AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781;

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

3) determining the methylation level of the one or more genes by polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and target capture.

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

1) measuring an amount of at least one methylated marker gene in DNA from the sample, wherein the one or more genes is selected from AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781;

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

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

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

1) measuring a methylation level of a CpG site for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with bisulfite a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent);

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

3) determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR;

wherein the one or more genes is selected from AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781.

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

Genomic DNA may be isolated by any means, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants, e.g., by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction, or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense, and required quantity of DNA. All clinical sample types comprising neoplastic matter or pre-neoplastic matter are suitable for use in the present method, e.g., cell lines, histological slides, biopsies, paraffin-embedded tissue, body fluids, stool, breast tissue, pancreatic tissue, leukocytes, colonic effluent, urine, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood, and combinations thereof.

The technology is not limited in the methods used to prepare the samples and provide a nucleic acid for testing. For example, in some embodiments, a DNA is isolated from a stool sample or from blood or from a plasma sample using direct gene capture, e.g., as detailed in U.S. Pat. Appl. Ser. No. 61/485386 or by a related method.

The genomic DNA sample is then treated with at least one reagent, or series of reagents, that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker comprising a DMR (e.g., DMR 1–13, as provided in Table 1).

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

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

The treated nucleic acid is then analyzed to determine the methylation state of the target gene sequences (at least one gene, genomic sequence, or nucleotide from a marker comprising a DMR, e.g., at least one DMR chosen from DMR 1–13, as provided in Table 1).

The method of analysis may be selected from those known in the art, including those listed herein, e.g., QuARTS and MSP as described herein.

Aberrant methylation, more specifically hypermethylation of a marker comprising a DMR (e.g., DMR 1–13, as provided in Table 1) is associated with PDAC.

The technology relates to the analysis of any sample associated with PDAC. For example, in some embodiments the sample comprises a tissue and/or biological fluid obtained from a patient. In some embodiments, the sample comprises a secretion. In some embodiments, the sample comprises blood, serum, plasma, gastric secretions, pancreatic juice, a gastrointestinal biopsy sample, microdissected cells from a breast biopsy, and/or cells recovered from stool. In some embodiments, the sample comprises pancreatic tissue. In some embodiments, the subject is human. The sample may include cells, secretions, or tissues from the endometrium, breast, liver, bile ducts, pancreas, stomach, colon, rectum, esophagus, small intestine, appendix, duodenum, polyps, gall bladder, anus, and/or peritoneum. In some embodiments, the sample comprises cellular fluid, ascites, urine, feces, pancreatic fluid, fluid obtained during endoscopy, blood, mucus, or saliva. In some embodiments, the sample is a stool sample. In some embodiments, the sample is a pancreatic tissue sample.

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

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

In some embodiments of the technology, a method for diagnosing PDAC in a subject is provided. The terms "diagnosing" and "diagnosis" as used herein refer to methods by which the skilled artisan can estimate and even determine whether or not a subject is suffering from a given disease or condition or may develop a given disease or condition in the future. The skilled artisan often makes a diagnosis on the basis of one or more diagnostic indicators, such as for example a biomarker (e.g., a DMR as disclosed herein), the methylation state of which is indicative of the presence, severity, or absence of the condition.

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

As such, "making a diagnosis" or "diagnosing", as used herein, is further inclusive of determining a risk of developing cancer or determining a prognosis, which can provide for predicting a clinical outcome (with or without medical treatment), selecting an appropriate treatment (or whether treatment would be effective), or monitoring a current treatment and potentially changing the treatment, based on the measure of the diagnostic biomarkers (e.g., DMR) disclosed herein. Further, in some embodiments of the presently disclosed subject matter, multiple determination of the biomarkers over time can be made to facilitate diagnosis and/or prognosis. A temporal change in the biomarker can be used to predict a clinical outcome, monitor the progression of PDAC, and/or monitor the efficacy of appropriate therapies directed against the cancer. In such an embodiment for example, one might expect to see a change in the methylation state of one or more biomarkers (e.g., DMR) disclosed herein (and potentially one or more additional biomarker(s), if monitored) in a biological sample over time during the course of an effective therapy.

The presently disclosed subject matter further provides in some embodiments a method for determining whether to initiate or continue prophylaxis or treatment of a cancer in a subject. In some embodiments, the method comprises providing a series of biological samples over a time period from the subject; analyzing the series of biological samples to determine a methylation state of at least one biomarker disclosed herein in each of the

biological samples; and comparing any measurable change in the methylation states of one or more of the biomarkers in each of the biological samples. Any changes in the methylation states of biomarkers over the time period can be used to predict risk of developing cancer, predict clinical outcome, determine whether to initiate or continue the prophylaxis or therapy of the cancer, and whether a current therapy is effectively treating the cancer. For example, a first time point can be selected prior to initiation of a treatment and a second time point can be selected at some time after initiation of the treatment. Methylation states can be measured in each of the samples taken from different time points and qualitative and/or quantitative differences noted. A change in the methylation states of the biomarker levels from the different samples can be correlated with PDAC risk, prognosis, determining treatment efficacy, and/or progression of the cancer in the subject.

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

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

As used herein, the phrase "determining the prognosis" refers to methods by which the skilled artisan can predict the course or outcome of a condition in a subject. The term "prognosis" does not refer to the ability to predict the course or outcome of a condition with 100% accuracy, or even that a given course or outcome is predictably more or less likely to occur based on the methylation state of a biomarker (e.g., a DMR). Instead, the skilled artisan will understand that the term "prognosis" refers to an increased probability that a certain

course or outcome will occur; that is, that a course or outcome is more likely to occur in a subject exhibiting a given condition, when compared to those individuals not exhibiting the condition. For example, in individuals not exhibiting the condition (e.g., having a normal methylation state of one or more DMR), the chance of a given outcome (e.g., suffering from PDAC) may be very low.

In some embodiments, a statistical analysis associates a prognostic indicator with a predisposition to an adverse outcome. For example, in some embodiments, a methylation state different from that in a normal control sample obtained from a patient who does not have a cancer can signal that a subject is more likely to suffer from a cancer than subjects with a level that is more similar to the methylation state in the control sample, as determined by a level of statistical significance. Additionally, a change in methylation state from a baseline (e.g., "normal") level can be reflective of subject prognosis, and the degree of change in methylation state can be related to the severity of adverse events. Statistical significance is often determined by comparing two or more populations and determining a confidence interval and/or a *p* value. See, e.g., Dowdy and Wearden, Statistics for Research, John Wiley & Sons, New York, 1983, incorporated herein by reference in its entirety. Exemplary confidence intervals of the present subject matter are 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% and 99.99%, while exemplary *p* values are 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, and 0.0001.

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

In some embodiments, a control sample is analyzed concurrently with the biological sample, such that the results obtained from the biological sample can be compared to the results obtained from the control sample. Additionally, it is contemplated that standard curves can be provided, with which assay results for the biological sample may be compared. Such standard curves present methylation states of a biomarker as a function of assay units, e.g., fluorescent signal intensity, if a fluorescent label is used. Using samples taken from multiple donors, standard curves can be provided for control methylation states of the one or more biomarkers in normal tissue, as well as for "at-risk" levels of the one or more biomarkers in tissue taken from donors with metaplasia or from donors with PDAC. In certain embodiments of the method, a subject is identified as having metaplasia upon identifying an aberrant methylation state of one or more DMR provided herein in a biological sample obtained from the subject. In other embodiments of the method, the detection of an aberrant methylation state of one or more of such biomarkers in a biological sample obtained from the subject results in the subject being identified as having cancer.

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

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

In some embodiments, the subject is diagnosed as having PDAC if, when compared to a control methylation state, there is a measurable difference in the methylation state of at least

one biomarker in the sample. Conversely, when no change in methylation state is identified in the biological sample, the subject can be identified as not having PDAC, not being at risk for the cancer, or as having a low risk of the cancer. In this regard, subjects having the cancer or risk thereof can be differentiated from subjects having low to substantially no cancer or risk thereof. Those subjects having a risk of developing PDAC can be placed on a more intensive and/or regular screening schedule, including endoscopic surveillance. On the other hand, those subjects having low to substantially no risk may avoid being subjected to additional testing for PDAC (e.g., invasive procedure), until such time as a future screening, for example, a screening conducted in accordance with the present technology, indicates that a risk of PDAC has appeared in those subjects.

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

Further with respect to diagnostic methods, a preferred subject is a vertebrate subject. A preferred vertebrate is warm-blooded; a preferred warm-blooded vertebrate is a mammal. A preferred mammal is most preferably a human. As used herein, the term “subject’ includes both human and animal subjects. Thus, veterinary therapeutic uses are provided herein. As such, the present technology provides for the diagnosis of mammals such as humans, as well as those mammals of importance due to being endangered, such as Siberian tigers; of economic importance, such as animals raised on farms for consumption by humans; and/or animals of social importance to humans, such as animals kept as pets or in zoos. Examples of

such animals include but are not limited to: carnivores such as cats and dogs; swine, including pigs, hogs, and wild boars; ruminants and/or ungulates such as cattle, oxen, sheep, giraffes, deer, goats, bison, and camels; and horses. Thus, also provided is the diagnosis and treatment of livestock, including, but not limited to, domesticated swine, ruminants, ungulates, horses (including race horses), and the like.

The presently-disclosed subject matter further includes a system for diagnosing PDAC in a subject. The system can be provided, for example, as a commercial kit that can be used to screen for a risk of PDAC or diagnose PDAC in a subject from whom a biological sample has been collected. An exemplary system provided in accordance with the present technology includes assessing the methylation state of a DMR as provided in Table 1.

EXAMPLES

Example I.

This example describes identification of plasma markers for detecting pancreatic ductal adenocarcinoma (PDAC).

13 methylated DNA markers (MDMs) were utilized in the identification of plasma markers for detecting pancreatic ductal adenocarcinoma (PDAC) (see, Table 1).

Table 1. Identified methylated regions distinguishing plasma from subjects having PDAC from plasma from subjects not having PDAC using the hg19 nomenclature.

DMR No.	Gene Annotation	Chromosome No.	Region on Chromosome (starting base-ending base)
1	AK055957	12	133484978-133485739
2	CD1D	1	158150797-158151205
3	CLEC11A	19	51228217-51228732
4	FER1L4	20	34189488-34189566
5	GRIN2D	19	48917755-48918477
6	HOXA1	7	27136145-27136425
7	LRRC4	7	127671993-127672310
8	MAX.chr5.4295	5	42951691-42951760
9	NTRK3	15	88800287-88800464
10	PRKCB	16	23846964-23848168
11	RYR2	1	237205577-237205684
12	SHISA9	16	12995930-12996219
13	ZNF781	19	38182950-38183127

The 13 MDMs shown in Table 1 originated from earlier pancreatic cancer tissue experiments using next generation bisulfite sequencing (see, Kisiel JB, et al., Clin Cancer Res. 2015 Oct 1;21(19):4473-81). Briefly, from re-mining these data hundreds of differentially methylated regions (DMRs) were identified based on a combination of selection criteria including area under the ROC curve (AUC), false discovery, relative and absolute % methylation difference between cases and controls, CpG density within the DMR, and (in cases) the presence of uniform contiguous co-methylation of neighboring residues. Subsequent validation using highly sensitive and specific targeted chemistries (quantitative methylation specific PCR, etc.) on larger sets of independent tissue samples allowed further marker refinement. Such selection yielded 20-30 potential MDMs, the majority of which mapped to putative or known regulatory regions – as determined from genome browser tracks. Many of the gene products operated in defined tumorigenic pathways and had functionality as promoter-associated transcription factors, enhancers, cell signaling mediators, growth factors, and ion channel proteins. Additional experiments were conducted to define a subset of very high performing discriminant assays (individual and complementary) which could be used in a formal plasma study. To this end, additional testing was performed on pools of neoplasia-free control plasma to eliminate MDMs which amplified from steady-state circulating cfDNA; an absolutely critical step. This resulted in the 13 MDMs shown in Table 1. Table 2 provides primer and probe information for the 13 MDMs recited in Table 1, and Fig. 1 further provides marker chromosomal regions used for the 13 MDMs recited in Table 1 and related primer and probe information.

Table 2.

DMR No.	Gene Annotation	Forward Primer	SEQ ID NO	Reverse Primer	SEQ ID NO	Probe	SEQ ID NO
1	AK055957	GATGGGTTTTAGAGGGGCGG	1	CGTACGACTCCCATTACCTTTAAACG	2	AGGCCACGGACG GCGACTCTCCGCCC/3C6/	3
2	CD1D	GGAGAAGAGTGCGTAGGTTAGAG	4	CATATCGCCCGACGTAAAAACC	5	CGCGCCGAGGCTCGCGAAACGCCG/3C6/	6
3	CLEC11A	GCGGGAGTTTGGCGTAG	7	CGCGCAAATACCGAATAAACG	8	CGCGCCGAGGGTCGGTAGATCGTTAGTAGATG/3C6/	9

4	FER1L4	CGTTGACGCGTAGTTTTCG	10	GTCGACCAAAAACGCGTC	11	CGCGCCGAGGCGTCCCGCAACTACAA/3C6/	12
5	GRIN2D	TCGATTATGTCGTTTTAGACGTTATCG	13	TCTACATCGACATTCTAAAACGACTAAC	14	AGGCCACGGACGCGCATACCATCGACTTCA/3C6/	15
6	HOXA1	AGTCGTTTTTTTAGGTAGTTTAGGCG	16	CGACCTTTACAATCGCCGC	17	CGCGCCGAGGGGCGGTAGTTGTTGC/3C6/	18
7	LRRC4	GCGTCGGCGTTAATTTCGC	19	ACAATACTCTTATATATTAACGCCGCTC	20	CGCGCCGAGGCGAGGTAGGCGACGG/3C6/	21
8	MAX.chr5.4295	GATTCGCGTTTTTTTTCGGATGGTC	22	TCTCGAATAAAAAAAACGACGCACG	23	AGGCCACGGACGCGATTAGACGGTTTTTTGTTAGT/3C6/	24
9	NTRK3	AGAGTTGGCGAGTTGGTTGTAC	25	CGAATTACAACAAAACCGAATAACGCGA	26	CGCGCCGAGGCGATACGGAAAGGCGT/3C6/	27
10	PRKCB	GTTGTTTTATATATCGGCGTTCGG	28	ACTACGACTATACACGCTTAACCG	29	CGCGCCGAGGGGTTATCGCGGGTTTCG/3C6/	30
11	RYR2	GGAGGTTTCGCGTTTCGATTA	31	CGAACGATCCCCGCCTAC	32	AGGCCACGGACGATTCGCGTTCGAGCG/3C6/	33
12	SHISA9	TGTTATGGGTTAGTGGGATTCGTC	34	CCGAAAACCACAAATCCCGC	35	CGCGCCGAGGCGTTTAATTGTAGTTCGGGC/3C6/	36
13	ZNF781	CGTTTTTTTGTTTTTCGAGTGCG	37	TCAATAACTAAACTCACCGCGTC	38	AGGCCACGGACGGCGGATTTATCGGGTTATAGT/3C6/	39

This panel of 13 MDMs were tested on a set of plasma samples from 26 patients diagnosed with PDAC (N=26; 4 S-I, 11 S-II, 6 S-III, 5 S-IV) and from normal EDTA plasma samples (N=26). Table 3 shows the area under the curve (AUC), fold change, p-value, percentage methylation for each marker. At 100% specificity, the 13 marker panel detected all of the stage 1 and stage 4 PDAC cancers, and all but one for each of the PDAC stage 2 and PDAC stage 3 cancers. In addition, this panel of 13 MDMs were tested on a set of PDAC tissue samples in comparison with benign tissue (Table 4), and were tested on a set of PDAC tissue samples in comparison with buffy coat (Table 5).

Table 3.

DMR No.	Gene Annotation	AUC	Fold Change	p-value	%Methylation PDAC plasma sample	%Methylation Control plasma sample

1	AK055957	0.84	26	<0.0001	2.045	0.078
2	CD1D	0.88	10	<0.0001	0.178	0.018
3	CLEC11A	0.82	60	<0.0001	1.481	.0.25
4	FER1L4	0.81	458	<0.0001	2.006	0.004
5	GRIN2D	0.79	24	<0.0001	0.498	0.020
6	HOXA1	0.83	49	<0.0001	0.645	0.013
7	LRRC4	0.80	7	<0.0001	10.415	1.548
8	MAX.chr5.4295	0.79	55	<0.0001	0.543	0.010
9	NTRK3	0.83	44	<0.0001	0.774	0.018
10	PRKCB	0.83	653	<0.0001	0.815	0.001
11	RYR2	0.70	15	0.0073	0.479	0.032
12	SHISA9	0.82	9	<0.0001	0.243	0.027
13	ZNF781	0.88	28	<0.0001	2.873	0.102

Table 4.

	PDAC vs benign tissue		
Marker	AUC	FC	pValue
AK055957	0.99	568	<0.0001
CD1D	1.00	>1000	<0.0001
CLEC11A	0.95	382	<0.0001
FER1L4	0.93	9	<0.0001
GRIN2D	0.95	6	<0.0001
HOXA1	0.89	18	<0.0001
LRRC4	0.91	0.41	<0.0001
MAX.chr5.4295	0.91	175	<0.0001
NTRK3	0.94	292	<0.0001
PRKCB	0.95	>1000	<0.0001
RYR2	0.98	81	<0.0001
SHISA9	0.95	168	<0.0001
ZNF781	0.95	>1000	<0.0001

Table 5.

	PDAC vs WBC		
Marker	AUC	FC	pValue
AK055957	0.99	>1000	<0.0001
CD1D	0.93	976	<0.0001
CLEC11A	0.93	988	<0.0001
FER1L4	1.00	604	<0.0001
GRIN2D	1.00	>1000	<0.0001
HOXA1	1.00	>1000	<0.0001
LRRC4	1.00	799	<0.0001
MAX.chr5.4295	0.94	>1000	<0.0001
NTRK3	1.00	391	<0.0001
PRKCB	0.93	>1000	<0.0001
RYR2	0.98	87	<0.0001
SHISA9	1.00	218	<0.0001

ZNF781	0.95	>1000	<0.0001

The only clinically available blood biomarker for detecting PDAC is CA 19-9. CA 19-9 is unreliable for early PDAC detection and may be normal in advanced disease. Experiments were next conducted to test the accuracy of the 13 markers shown in Table 1 with or without CA19-9 to discriminate PDAC cases from age-sex balanced control patients.

All assays with the 13 markers were performed in blinded fashion by target enrichment long-probe quantitative amplified signal (TELQAS) testing (see, Kisiel JB, et al., Hepatology. 2018 Aug 31). Briefly, TELQAS oligos (forward invasive primer, reverse primer, flap probe) were designed to CpG motifs within each of the 13 DMRs (IDT, Coralville IA). 12 cycles of multiplex amplification of the markers as well as B3GALT6 (reference gene) and RASSF1 (zebrafish processing control) were performed. The products were then diluted 10-fold with TE buffer; 10 μL of the diluted amplicons were used in triplex format (FAM, HEX, Quasar 670) in which two markers plus the B3GALT6 reference gene were amplified and quantified. TELQAS reactions were performed on ABI 7500DX equipment (Applied Biosystems, Foster City CA).

CA 19-9 was quantitated from plasma samples using the MILLIPLEX® Map Kit (EMD Millipore) on the Luminex® MAGPIX® analyzer. Briefly, plasma samples were diluted 1:6 using the Serum Matrix provided in the kit as the diluent. Only CA19-9 Antibody-Immobilized Magnetic Beads were used in the immunoassay. The assay was completed using the protocol supplied with the kit reagents. Quantitative results for each sample were generated from the median fluorescence intensity signals using the Luminex® xPONENT® software.

From 340 plasma samples (170 PDAC cases, 170 controls) experiments initially used quantitative MDM and CA19-9 levels in 120 advanced stage PDAC cases (60 Stage 3 and 60 Stage 4) and 120 healthy controls to train a prediction algorithm by random forest (rForest) modeling at 97.5% specificity. A locked algorithm was then applied to an independent blinded test set of 50 early stage PDAC cases (5 Stage 1, 45 Stage 2) and 50 controls. Subsequently, data from all 340 patients were combined and refit using rForest. The MDM panel was cross-validated by randomly splitting the entire data set 2:1 for training and testing. The fitted rForest model from the training set was used to predict disease status in the testing set; median AUCs were reported after 500 iterations.

Area under the curve results for the 13 markers is shown in Table 6. In the initial training set, the MDM-CA19-9 panel detected 54/60 (90%) Stage 3 and 59/60 (98%) Stage 4

PDACs at 97.5% specificity. Area under the curve MDM cut-off values derived from these advanced-stage cases and applied to stage 1 & 2 PDAC and controls yielded an AUC of 0.84 (95% CI 0.76-0.92) by MDM panel alone vs 0.91 (0.84-0.97) by combined MDM-CA19-9 panel (p=0.038). Combining all 340 cases and controls, the cross-validated sensitivity of the MDM-CA 19-9 panel was 79% in Stage 1, 82% in Stage 2, 94% in Stage 3 and 99% in Stage 4 PDAC at a specificity of 92% (81-100%) (Figure 2). The cross-validated AUC was 0.9 (0.85−0.94) for the MDM panel alone vs 0.97 (0.94−0.99) for the combined MDM-CA 19-9 panel, p=<0.0001 (Figure 3). Overall, sensitivity for PDAC was 92% (83-98%) at 92% specificity. Such results indicate that the 13 MDMs shown Table 1 in combination or not in combination with CA19-9 detect PDAC across all stages with moderate to high accuracy.

Table 6.

	ROC AUC	lower.95	upper.95
GRIN2D	0.76	0.71	0.81
CD1D	0.80	0.75	0.85
ZNF781	0.78	0.73	0.83
FER1L4	0.81	0.77	0.86
RYR2_F	0.76	0.70	0.81
CLEC11A	0.81	0.77	0.86
MAX_Chr12_1334	0.80	0.75	0.84
LRRC4	0.75	0.70	0.80
MAX_Chr5_4295	0.83	0.78	0.87
HOXA1	0.84	0.80	0.88
PRKCB	0.82	0.78	0.87
SHISA9	0.81	0.76	0.86
NTRK3	0.75	0.69	0.80

10cc of blood from each subject was collected in a K2EDTA vacutainer (BD, Franklin Lakes NJ). Within 4 hours, the tubes were centrifuged at 1500xG (10 min), plasma removed and centrifuged a second time, aliquoted in 2mL cryotubes, and stored at -80°C without any intermittent thawing. The cfDNA was purified and bisulfite converted using an automated silica bead method. A non-human DNA spike was used to control for processing aberrations. For all samples, 3.8 mL of plasma was initially subjected to Proteinase K treatment followed by lysis with detergent and chaotropic reagents. Silica coated binding

beads and lysis buffer containing isopropyl alcohol were added to each sample for DNA capture and DNA precipitation. All samples were subjected to multiple rounds of washing on the Hamilton STARlet liquid handling system (Hamilton Company, Reno NV) and binding beads were dried prior to DNA sample elution in elution buffer. Samples were then bisulfite converted as previously described (see, Lidgard, et al., 2013;11:1313-1318) with the use of the Hamilton STARlet liquid handling system. Briefly, samples were initially denatured with sodium hydroxide. Ammonium bisulfite was added to each sample for deamination. Samples were subsequently bound to silica coated binding beads and subjected to multiple rounds of washing prior to desulphonation. Sample washing was repeated, and purified samples were eluted in elution buffer.

The sample cfDNA was tested using TELQAS (target enrichment with long probe quantitative amplified signal), a highly sensitive multiplexed assay format. (see, Kisiel JB, et al., Hepatology. 2018 Aug 31). Briefly, TELQAS oligos (forward invasive primer, reverse primer, flap probe) were designed to CpG motifs within each of the 13 DMRs (IDT, Coralville IA). 12 cycles of multiplex amplification of the markers as well as B3GALT6 (reference gene) and RASSF1 (zebrafish processing control) were performed. The products were then diluted 10-fold with TE buffer; 10 µL of the diluted amplicons were used in triplex format (FAM, HEX, Quasar 670) in which two markers plus the B3GALT6 reference gene were amplified and quantified. TELQAS reactions were performed on ABI 7500DX equipment (Applied Biosystems, Foster City CA). Table 7 shows the 9 LQAS assays that were run. All LQAS assays were setup and run with standard, previously published conditions.

Table 7. 9 Biplex/Triplex marker configurations

#	A5	A1	A3	Biplex/Triplex
1	ZF-RASSF1_wt	-	B3GALT6_wt	ZB_WT
2	GRIN2D	CD1D	B3GALT6	GCB
3	ZNF781	FER1L4	B3GALT6	ZFB
4	RYR2_F	CLEC11A	B3GALT6	RCB
5	MAX.Chr12.1334	LRRC4	B3GALT6	MLB
6	MAX.Chr5.4295	HOXA1	B3GALT6	MHB
7	-	PRKCB	B3GALT6	PB
8	-	SHISA9	B3GALT6	SB
9	ZF-RASSF1	NTRK3	B3GALT6	ZNB

Example II.

The testing of the panel of 13 MDMs recited in Table 1 were further tested on a collection of LBgard (Biomatrica, San Diego, CA) plasma samples comprised of 12 patients diagnosed with PDAC (N=12; 3 S-II, 1 S-III, 8 S-IV) and from 27 normal non-PDAC plasma samples (N=27). Table 8 shows the nominal logistic fit for assessing if the sample is from a control or a PDAC case.

Table 8.

Nominal Logistic Fit for Is the patient a control or case

Effect Summary

Source	LogWorth	PValue
%SHISA9	99.492	0.00000
%NTRK3	88.819	0.00000
%FER1L4	41.620	0.00000
%CLEC11A	32.875	0.00000
%MAX.Chr5.4295	6.191	0.00000
%PRKCB	4.717	0.00002
%HOXA1	0.045	0.90192
%GRIN2D	0.003	0.99365
%LRRC4	0.002	0.99574
%RYR2_F	0.001	0.99786
%CD1D	0.000	0.99933
%ZNF781	0.000	0.99951
%MAX.Chr12.1334	0.000	1.00000

Having now fully described the invention, it will be understood by those of skill in the art that the same can be performed within a wide and equivalent range of conditions, formulations, and other parameters without affecting the scope of the invention or any embodiment thereof. All patents, patent applications and publications cited herein are fully incorporated by reference herein in their entirety.

INCORPORATION BY REFERENCE

The entire disclosure of each of the patent documents and scientific articles referred to herein is incorporated by reference for all purposes.

EQUIVALENTS

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are intended to be embraced therein.

CLAIMS

WE CLAIM:

1. A method, comprising:

measuring a methylation level for one or more genes in a biological sample of a human individual through

treating genomic DNA in the biological sample with a reagent that modifies DNA in a methylation-specific manner;

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

determining the methylation level of the one or more genes by polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and target capture;

wherein the one or more genes is selected from AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781.

2. The method of claim 1, wherein the DNA is treated with a reagent that modifies DNA in a methylation-specific manner.

3. The method of claim 2, wherein the reagent comprises one or more of a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent.

4. The method of claim 3, wherein the DNA is treated with a bisulfite reagent to produce bisulfite-treated DNA.

5. The method of claim 1, wherein the measuring comprises multiplex amplification.

6. The method of claim 1, wherein measuring the amount of at least one methylated marker gene comprises using one or more methods selected from the group consisting of methylation-specific PCR, quantitative methylation-specific PCR, methylation-specific DNA

restriction enzyme analysis, quantitative bisulfite pyrosequencing, flap endonuclease assay, PCR-flap assay, and bisulfite genomic sequencing PCR.

7. The method of claim 1, wherein the sample comprises one or more of a plasma sample, a blood sample, or a tissue sample (e.g., pancreatic tissue).

8. The method of claim 1, wherein the set of primers for the selected one or more genes is recited in Table 2.

9. A method of characterizing a sample, comprising:

a) measuring an amount of at least one methylated marker gene in DNA from the sample, wherein the at least one methylated marker gene is one or more genes selected from AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781;

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

c) calculating a value for the amount of the at least one methylated marker gene measured in the DNA as a percentage of the amount of the reference marker gene measured in the DNA, wherein the value indicates the amount of the at least one methylated marker DNA measured in the sample.

10. The method of claim 9, wherein the at least one reference marker comprises one or more reference marker selected from *B3GALT6* DNA and β-actin DNA.

11. The method of claim 9, wherein the sample comprises one or more of a plasma sample, a blood sample, or a tissue sample (e.g., pancreatic tissue).

12. The method of claim 9, wherein the DNA is extracted from the sample.

13. The method of claim 9, wherein the DNA is treated with a reagent that modifies DNA in a methylation-specific manner.

14. The method of claim 13, wherein the reagent comprises one or more of a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent.

15. The method of claim 14 wherein the DNA is treated with a bisulfite reagent to produce bisulfite-treated DNA.

16. The method of claim 14, wherein the modified DNA is amplified using a set of primers for the selected one or more genes.

17. The method of claim 16, wherein the set of primers for the selected one or more genes is recited in Table 2.

18. The method of claim 9 wherein measuring amounts of a methylated marker gene comprises using one or more of polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and target capture.

19. The method of claim 18, wherein the measuring comprises multiplex amplification.

20. The method of claim 18, wherein measuring the amount of at least one methylated marker gene comprises using one or more methods selected from the group consisting of methylation-specific PCR, quantitative methylation-specific PCR, methylation-specific DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, flap endonuclease assay, PCR-flap assay, and bisulfite genomic sequencing PCR.

21. A method for characterizing a biological sample comprising:

(a) measuring a methylation level of a CpG site for one or more genes selected from AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 in a biological sample of a human individual through

treating genomic DNA in the biological sample with bisulfite;

amplifying the bisulfite-treated genomic DNA using a set of primers for the selected one or more genes; and

determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR;

(b) comparing the methylation level to a methylation level of a corresponding set of genes in control samples without PDAC; and

(c) determining that the individual has PDAC when the methylation level measured in the one or more genes is higher than the methylation level measured in the respective control samples.

22. The method of claim 21 wherein the set of primers for the selected one or more genes is recited in Table 2.

23. The method of claim 21, wherein the biological sample is a plasma sample, a blood sample, or a tissue sample (e.g., pancreatic tissue).

24. The method of claim 21, wherein the one or more genes is described by the genomic coordinates shown in Table 1.

25. The method of claim 21, wherein said CpG site is present in a coding region or a regulatory region.

26. The method of claim 21, wherein said measuring the methylation level a CpG site for one or more genes comprises a determination selected from the group consisting of determining the methylation score of said CpG site and determining the methylation frequency of said CpG site.

27. A method, comprising:

(a) measuring a methylation level of a CpG site for one or more genes selected from AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781 in a biological sample of a human individual through

treating genomic DNA in the biological sample with bisulfite;

amplifying the bisulfite-treated genomic DNA using a set of primers for the selected one or more genes; and

determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR.

28. The method of claim 27 wherein the set of primers for the selected one or more genes is recited in Table 2.

29. The method of claim 27, wherein the biological sample is a plasma sample, a blood sample, or a tissue sample (e.g., pancreatic tissue).

30. The method of claim 27, wherein the one or more genes is described by the genomic coordinates shown in Table 1.

31. The method of claim 27, wherein said measuring the methylation level a CpG site for one or more genes comprises a determination selected from the group consisting of determining the methylation score of said CpG site and determining the methylation frequency of said CpG site.

32. A method of screening for PDAC in a sample obtained from a subject, the method comprising:

1) assaying a methylation state of a DNA methylation marker comprising a chromosomal region having an annotation selected from the group consisting of AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781, and

2) identifying the subject as having PDAC when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have PDAC.

33. The method of claim 32 comprising assaying a plurality of markers.

34. The method of claim 32 wherein the marker is in a high CpG density promoter.

35. The method of claim 32 wherein the sample is a stool sample, a tissue sample, a pancreatic tissue sample, a plasma sample, or a urine sample.

36. The method of claim 32 wherein the assaying comprises using methylation specific polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation specific nuclease, mass-based separation, or target capture.

37. The method of claim 32 wherein the assaying comprises use of a methylation specific oligonucleotide.

38. A method for characterizing a sample from a human patient comprising:

a) obtaining DNA from a sample of a human patient;

b) assaying a methylation state of a DNA methylation marker comprising a chromosomal region having an annotation selected from the group consisting of AK055957, CD1D, CLEC11A, FER1L4, GRIN2D, HOXA1, LRRC4, MAX.chr5.4295, NTRK3, PRKCB, RYR2, SHISA9, and ZNF781;

c) comparing the assayed methylation state of the one or more DNA methylation markers with methylation level references for the one or more DNA methylation markers for human patients not having PDAC.

39. The method of claim 38 wherein the sample is a stool sample, a tissue sample, a pancreatic tissue sample, a plasma sample, or a urine sample.

40. The method of claim 38 comprising assaying a plurality of DNA methylation markers.

41. The method of claim 38 wherein the assaying comprises using methylation specific polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation specific nuclease, mass-based separation, or target capture.

42. The method of claim 38 wherein the assaying comprises use of a methylation specific oligonucleotide.

43. The method of claim 38 wherein the methylation specific oligonucleotide is selected from a set of primers for the selected one or more genes is recited in Table 2 or a probe selected from Table 2.

44. A method for characterizing a sample obtained from a human subject, the method comprising reacting a nucleic acid comprising a DMR with a bisulfite reagent to produce a bisulfite-reacted nucleic acid; sequencing the bisulfite-reacted nucleic acid to provide a nucleotide sequence of the bisulfite-reacted nucleic acid; comparing the nucleotide sequence of the bisulfite-reacted nucleic acid with a nucleotide sequence of a nucleic acid comprising the DMR from a subject who does not have PDAC to identify differences in the two sequences.

45. A system for characterizing a sample obtained from a human subject, the system comprising an analysis component configured to determine the methylation state of a sample, a software component configured to compare the methylation state of the sample with a control sample or a reference sample methylation state recorded in a database, and an alert component configured to determine a single value based on a combination of methylation states and alert a user of a PDAC-associated methylation state.

46. The system of claim 45 wherein the sample comprises a nucleic acid comprising a DMR.

47. The system of claim 45 further comprising a component for isolating a nucleic acid.

48. The system of claim 45 further comprising a component for collecting a sample.

49. The system of claim 45 wherein the sample is a stool sample, a tissue sample, a pancreatic tissue sample, a plasma sample, or a urine sample.

50. The system of claim 45 wherein the database comprises nucleic acid sequences comprising a DMR.

51. The system of claim 45 wherein the database comprises nucleic acid sequences from subjects who do not have PDAC.

52. A kit comprising:
1) a bisulfite reagent; and
2) a control nucleic acid comprising a sequence from a DMR selected from a group consisting of DMR 1–13 from Table 1, and having a methylation state associated with a subject who does not have PDAC.

53. A kit comprising a bisulfite reagent and an oligonucleotide according to SEQ ID NOS 1-39.

54. A kit comprising a sample collector for obtaining a sample from a subject; reagents for isolating a nucleic acid from the sample; a bisulfite reagent; and an oligonucleotide according to SEQ ID NOS 1-39.

55. The kit according to claim 53 wherein the sample is a stool sample, a tissue sample, a pancreatic tissue sample, a plasma sample, or a urine sample.

56. A composition comprising a nucleic acid comprising a DMR and a bisulfite reagent.

57. A composition comprising a nucleic acid comprising a DMR and an oligonucleotide according to SEQ ID NOS 1-39.

58. A composition comprising a nucleic acid comprising a DMR and a methylation-sensitive restriction enzyme.

59. A composition comprising a nucleic acid comprising a DMR and a polymerase.

60. A method for screening for PDAC in a sample obtained from a subject, the method comprising reacting a nucleic acid comprising a DMR with a bisulfite reagent to produce a bisulfite-reacted nucleic acid; sequencing the bisulfite-reacted nucleic acid to provide a nucleotide sequence of the bisulfite-reacted nucleic acid; comparing the nucleotide sequence of the bisulfite-reacted nucleic acid with a nucleotide sequence of a nucleic acid comprising the DMR from a subject who does not have PDAC to identify differences in the two sequences; and identifying the subject as having PDAC when a difference is present.

61. A system for screening for PDAC in a sample obtained from a subject, the system comprising an analysis component configured to determine the methylation state of a sample, a software component configured to compare the methylation state of the sample with a control sample or a reference sample methylation state recorded in a database, and an alert component configured to determine a single value based on a combination of methylation states and alert a user of a PDAC-associated methylation state.

62. The system of claim 61 wherein the sample comprises a nucleic acid comprising a DNA methylation marker comprising a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1–13 from Table 1.

63. The system of claim 61 further comprising a component for isolating a nucleic acid.

64. The system of claim 61 further comprising a component for collecting a sample.

65. The system of claim 61 further comprising a component for collecting a stool sample, a pancreatic tissue sample, and/or a plasma sample.

66. The system of claim 61 wherein the database comprises nucleic acid sequences from subjects who do not have PDAC.

67. The method of claim 1, further comprising

measuring the level of carbohydrate antigen 19-9 (CA19-9) from the biological sample.

68. The method of claim 21, further comprising

measuring the level of carbohydrate antigen 19-9 (CA19-9) from the biological sample;

comparing the measured level of CA19-9 with a reference level for CA19-9 from a control biological sample; and

determining that the individual has PDAC when the methylation level measured in the one or more genes is higher than the methylation level measured in the respective control samples and the level of CA19-9 is higher than the reference level for CA19-9 from the control biological sample.

69. The method of claim 27, further comprising

measuring the level of carbohydrate antigen 19-9 (CA19-9) from the biological sample.

FIG. 1

CD1D
>hg19_dna range=chr1:158150818 158150915 strand=+

WT:
CGGCGGGGGAGAAGAGTGCGCAGGTCAGAGGGCGGCGCGCAGCGGCGCTCCGCGAGGTCCCCACGCCGGGCGATATGGGGTGCCTGCTGTTTCT
GCT (SEQ ID NO:40)
G
BST:
CGGCGGGGGAGAAGAGTGCGTAGGTTAGAGGCGGCGCGTAGCGGCGTTTCGCGAGGTTTTTACGTCGGGCGATATGGGGTGTTTGTTGTTTT
TGTTG (SEQ ID NO:41)

CD1D_FP_LQ GGAGAAGAGTGCGTAGGTTAGAG (SEQ ID NO:4)
CD1D_RP_2a CATATCGCCCGACGTAAAAACC (SEQ ID NO:5)
CD1D_LQ_Pb_A1 CGCGCCGAGG CTCGCGAAACGCCG/3C6/ (SEQ ID NO:6)

FER1L4
>hg19_dna range=chr20:34189490-34189607 strand=-
WT:
CCCGAATGGAACGAGCAGCTGAGCTTCGTGGAGCTCTTCCCGCCGCTGACGCGCAGCCTCCGCCTGCAGCTGCGGGACGACGCGCCCCTGGTCG
ACGCGGCACTCGCTACGCACGTGC (SEQ ID NO:42)

BST:
TTCGAATGGAACGAGTAGTTGAGTTTCGTGGAGTTTTTTTCGTCGTTGACGCGTAGTTTTCGTTTGTAGTTGCGGGACGACGCGTTTTTGGTC
GACGCGGTATTCGTTACGTACGTGT (SEQ ID NO:43)

FER1L4_FP CGTTGACGCGTAGTTTTCG (SEQ ID NO:10)
FER1L4_RP GTCGACCAAAAACGCGTC (SEQ ID NO:11)
FER1L4_LQ_Pb_A1 CGCGCCGAGG CGTCCCGCAACTACAA/3C6/ (SEQ ID NO:12)

CLEC11A
>hg19_dna range=chr19:51228314-51228507 5'pad=0 3'pad=0 strand=+
WT:
TTCCTGCTCTCGCGCGACTTCGAAGCTCAGGCGGCGGCGCAGGCGCGGTGCACGGCGCGGGGCGGGAGCCTGGCGCAGCCGGCAGACCGCCAGC
AGATGGAGGCGCTCACTCGGTACCTGCGCGCGGCGCTCGCTCCCTACAACTGGCCCGTGTGGCTGGGCGTGCACGATCGGCGCGCCGAGGGCCT
CTACCT (SEQ ID NO:44)

BST:
TTTTTGTTTTCGCGCGATTTCGAAGTTTAGGCGGCGGCGTAGGCGCGGTGTACGGCGCGGGGCGGGAGTTTGGCGTAGTCGGTAGATCGTTAG
TAGATGGAGGCGTTTATTCGGTATTTGCGCGCGGCGTTCGTTTTTTATAATTGGTTCGTGTGGTTGGGCGTGTACGATCGGCGCGTCGAGGGT
TTTTATTT (SEQ ID NO:45)

CLEC_FP_1 GCGGGAGTTTGGCGTAG (SEQ ID NO:7)
CLEC_RP_1 CGCGCAAATACCGAATAAACG (SEQ ID NO:8)
CLEC11A_Pb_A1_63 CGCGCCGAGG GTCGGTAGATCGTTAGTAGATG/3C6/ (SEQ ID NO:9)

LRRC4
>hg19_dna range=chr7:127672186-127672282 5'pad=0 3'pad=0 strand=-

WT:
GGCGCGGCGCTGGAAGGCGCCGGCGTTAACCCCGCGAGGCAGGCGACGGAGGGGGAGCGGCGCTAATACATAAGAGCACTGCATCACGCTAATC
TTC (SEQ ID NO:46)

BST:
GGCGCGGCGTTGGAAGGCGTCGGCGTTAATTTCGCGAGGTAGGCGACGGAGGGGGAGCGGCGTTAATATATAAGAGTATTGTATTACGTTAAT
TTTT (SEQ ID NO:47)

LRRC4_FP_v2 GCGTCGGCGTTAATTTCGC (SEQ ID NO:19)
LRRC4_RP_v2 ACAATACTCTTATATATTAACGCCGCTC (SEQ ID NO:20)
LRRC4_LQ_Pb_A1 CGCGCCGAGG CGAGGTAGGCGACGG/3C6/ (SEQ ID NO:21)

HOXA1
>hg19_dna range=chr7:27135789-27135861 5'pad=0 3'pad=0 strand=+
WT:
GACGCGCAGTCGCCCCCCCAGGCAGCCTAGGCGGCGGCAGCTGCTGCGGCGACTGCAAAGGCCGATTTGGAGT (SEQ ID NO:48)

BST:
GACGCGTAGTCGTTTTTTTAGGTAGTTTAGGCGGCGGTAGTTGTTGCGGCGATTGTAAAGGTCGATTTGGAGT (SEQ ID NO:49)

HOXA1_FP AGTCGTTTTTTTAGGTAGTTTAGGCG (SEQ ID NO:16)

FIG. 1 (CONT'D)

HOXA1_RP CGACCTTTACAATCGCCGC (SEQ ID NO:17)
HOXA1_Pb_A1_63 CGCGCCGAGG GGCGGTAGTTGTTGC/3C6/ (SEQ ID NO:18)

PRKCB: Chr16:23847120-23847216 strand=-
WT:
AGGCCCCGCCCAGCGCTGCCAGCTGCTTTACATATCGGCGCCCGGGCTACCGCGGGCCTCGCGGCTAAGCGTGCACAGCCGCAGCTCTGCAGCG
CCG (SEQ ID NO:50)

BST:
AGGTTTCGTTTAGCGTTGTTAGTTGTTTTATATATCGGCGTTCGGGTTATCGCGGGTTTCGCGGTTAAGCGTGTATAGTCGTAGTTTTGTAGC
GTCG (SEQ ID NO:51)

PRKCB_FP_1 GTTGTTTTATATATCGGCGTTCGG (SEQ ID NO:28)
PRKCB_RP_1_LQ ACTACGACTATACACGCTTAACCG (SEQ ID NO:29)
PRCKB_LQ_Pb_A1 CGCGCCGAGG GGTTATCGCGGGTTTCG/3C6/ (SEQ ID NO:30)

SHISA9:
>hg19_dna range=Chr16:12996156-12996250 5'pad=0 3'pad=0 strand=+
WT:
GATGTCATGGGCCAGTGGGACCCGCCGTTCAACTGCAGCTCGGGCGACTTCATCTTCTGCTGCGGGACTTGTGGCTTCCGGTTCTGCTGCACGT
T (SEQ ID NO:52)

BST:
GATGTTATGGGTTAGTGGGATTCGTCGTTTAATTGTAGTTCGGGCGATTTTATTTTTTGTTGCGGGATTTGTGGTTTTCGGTTTTGTTGTACG
TT (SEQ ID NO:53)

SHISA9_FP TGTTATGGGTTAGTGGGATTCGTC (SEQ ID NO:34)
SHISA9_RP CCGAAAACCACAAATCCCGC (SEQ ID NO:35)
SHISA9_LQ_Pb_A1 CGCGCCGAGG CGTTTAATTGTAGTTCGGGC/3C6/ (SEQ ID NO:36)

NTRK3
>hg19_dna range=Chr15:88800303-88800464 5'pad=0 3'pad=0 strand=+
WT:
CGAAGGGAAAAAGTTGCATTTGAGATTGCGAGGGTCGGGGCTGGGGAGGGAGAGTTGGCGAGCTGGCTGCACGACACGGAAAGGCGCTCTCCTT
TCCACTTTTTGGCCCTCGCGCTACCCGGTTTTGCTGCAATCCGGACCGCGGTAGGAAGTGAAATGAAC (SEQ ID NO:54)

BST:
CGAAGGGAAAAAGTTGTATTTGAGATTGCGAGGGTCGGGGTTGGGGAGGGAGAGTTGGCGAGTTGGTTGTACGATACGGAAAGGCGTTTTTTT
TTTTATTTTTTGGTTTTCGCGTTATTCGGTTTTGTTGTAATTCGGATCGCGGTAGGAAGTGAAATGAAT (SEQ ID NO:55)

NTRK3_FP AGAGTTGGCGAGTTGGTTGTAC (SEQ ID NO:25)
NTRK3_RP CGAATTACAACAAAACCGAATAACGCGA (SEQ ID NO:26)
NTRK3_LQ_Pb_A1 CGCGCCGAGG CGATACGGAAAGGCGT/3C6/ (SEQ ID NO:27)

GRIN2D
>hg19_dna range=chr19:48918160-48918300 strand=-
WT:
CGCCCCCTCACCTCCCCGATCATGCCGTTCCAGACGCCATCGATCTTCTTTCCGTGCTTGCCATTGGTGACCAGGTAGAGGTCGTAGCTGAAGC
CGATGGTATGCGCCAGCCGCTTCAGAATGTCGATGCAGAAACCCTTG (SEQ ID NO:56)
BST:
CGTTTTTTTATTTTTTCGATTATGTCGTTTTAGACGTTATCGATTTTTTTTTTCGTGTTTGTTATTGGTGATTAGGTAGAGGTCGTAGTTGAAG
TCGATGGTATGCGTTAGTCGTTTTAGAATGTCGATGTAGAAATTTTTG (SEQ ID NO:57)

GRIN2D FP TCGATTATGTCGTTTTAGACGTTATCG (SEQ ID NO:13)
GRIN2D RP TCTACATCGACATTCTAAAACGACTAAC (SEQ ID NO:14)
GRIN2D_Pb_A5_LQ AGGCCACGGACG CGCATACCATCGACTTCA/3C6/ (SEQ ID NO:15)

ZNF781:Chr19: 38183018-38183137 strand=- (GRCh37/hg19)
WT:
AAGCTGCGCCCGGAGACGTGGGAGCGTTCTCTTGTTTTCCGAGTGCGCGGACTCATCGGGTCACAGTTTATGCTTTTATGACGCGGTGAGTCCA
GCCACTGATTCCTAACGGTTTAGAGT (SEQ ID NO:58)

BST:
AAGTTGCGTTCGGAGACGTGGGAGCGTTTTTTTGTTTTTCGAGTGCGCGGATTTATCGGGTTATAGTTTATGTTTTTATGACGCGGTGAGTTT
AGTTATTGATTTTTAACGGTTTAGAGT (SEQ ID NO:59)

FIG. 1 (CONT'D)

ZNF781 F.primer CGTTTTTTTGTTTTTCGAGTGCG (SEQ ID NO:37)
ZNF781 R.primer TCAATAACTAAACTCACCGCGTC (SEQ ID NO:38)
ZNF781_Pb_A5_63_v2 AGGCCACGGACG GCGGATTTATCGGGTTATAGT/3C6/ (SEQ ID NO:39)

RYR2_F: chr1:237205546-237205717 strand=+
WT:
TGCGGGGCTGCTTCCCCGCGTCCTCCGGGCCCGGGCCGCCCTCCTCCCGCACAGTGCGGAGCAGGGAGGCCCCGCGCCTCGACCACCCGCGCCC
GAGCGTCCGCGCCTCCTCCTCCGCTCTGCAGGCGGGGACCGCCCGGCGCTCGGCACCCGGCAGCGCGGCCCCCTCCAG (SEQ ID NO:60)

BST:
TGCGGGGTTGTTTTTTCGCGTTTTTCGGGTTCGGGTCGTTTTTTTTTTCGTATAGTGCGGAGTAGGGAGGTTTCGCGTTTCGATTATTCGCGTT
CGAGCGTTCGCGTTTTTTTTTTCGTTTTGTAGGCGGGGATCGTTCGGCGTTCGGTATTCGGTAGCGCGGTTTTTTTTAG (SEQ ID NO:61)

RYR2_F_FP_v2 GGAGGTTTCGCGTTTCGATTA (SEQ ID NO:31)

RYR2_F_RP_v2 CGAACGATCCCCGCCTAC (SEQ ID NO:32)
RYR2_F_LQ_Pb_A5 AGGCCACGGACG ATTCGCGTTCGAGCG/3C6/ (SEQ ID NO:33)

MAX.Chr12.1334
>hg19_dna range=Chr12:133485074-133485156 5'pad=0 3'pad=0 strand=+
WT:
GGATGGGTCCCAGAGGGGCGGGTCGGGGCGGAGAGCCGCGCCCAAAGGCAATGGGAGCCGCACGCTGCTAGGCAACATGCTGT (SEQ ID NO:62)

BST:
GGATGGGTTTTAGAGGGGCGGGTCGGGGCGGAGAGTCGCGTTTAAAGGTAATGGGAGTCGTACGTTGTTAGGTAATATGTTGT (SEQ ID NO:63)

MAX.Chr12.1334_FP GATGGGTTTTAGAGGGGCGG (SEQ ID NO:1)
MAX.Chr12.1334_RP CGTACGACTCCCATTACCTTTAAACG (SEQ ID NO:2)
MAX.Chr12.1334_LQ_Pb_A5 AGGCCACGGACG GCGACTCTCCGCCC/3C6/ (SEQ ID NO:3)

MAX.Chr5.4295
>hg19_dna range=Chr5:42951661-42951790 5'pad=0 3'pad=0 strand=-
WT:
TTTCTGATCCGCGCTTTCTCCCGGATGGCCGACCAGACGGTTTTCTGCTAGTTGTGCTACACATAGTTTCCTGGTCTCCGTGCGCCGCTTTCTC
CATCCGAGACCCTCTAGTGCCGGTGTCACGGTCGCA (SEQ ID NO:64)

BST:
TTTTTGATTCGCGTTTTTTTTCGGATGGTCGATTAGACGGTTTTTTGTTAGTTGTGTTATATATAGTTTTTTGGTTTTCGTGCGTCGTTTTTT
TTATTCGAGATTTTTTAGTGTCGGTGTTACGGTCGTA (SEQ ID NO:65)

MAX.Chr5.4295_FP GATTCGCGTTTTTTTTCGGATGGTC (SEQ ID NO:22)
MAX.Chr5.4295_RP TCTCGAATAAAAAAAACGACGCACG (SEQ ID NO:23)
MAX.Chr5.4295_LQ_Pb_A5 AGGCCACGGACG CGATTAGACGGTTTTTTGTTAGT/3C6/ (SEQ ID NO:24)

FIG. 2

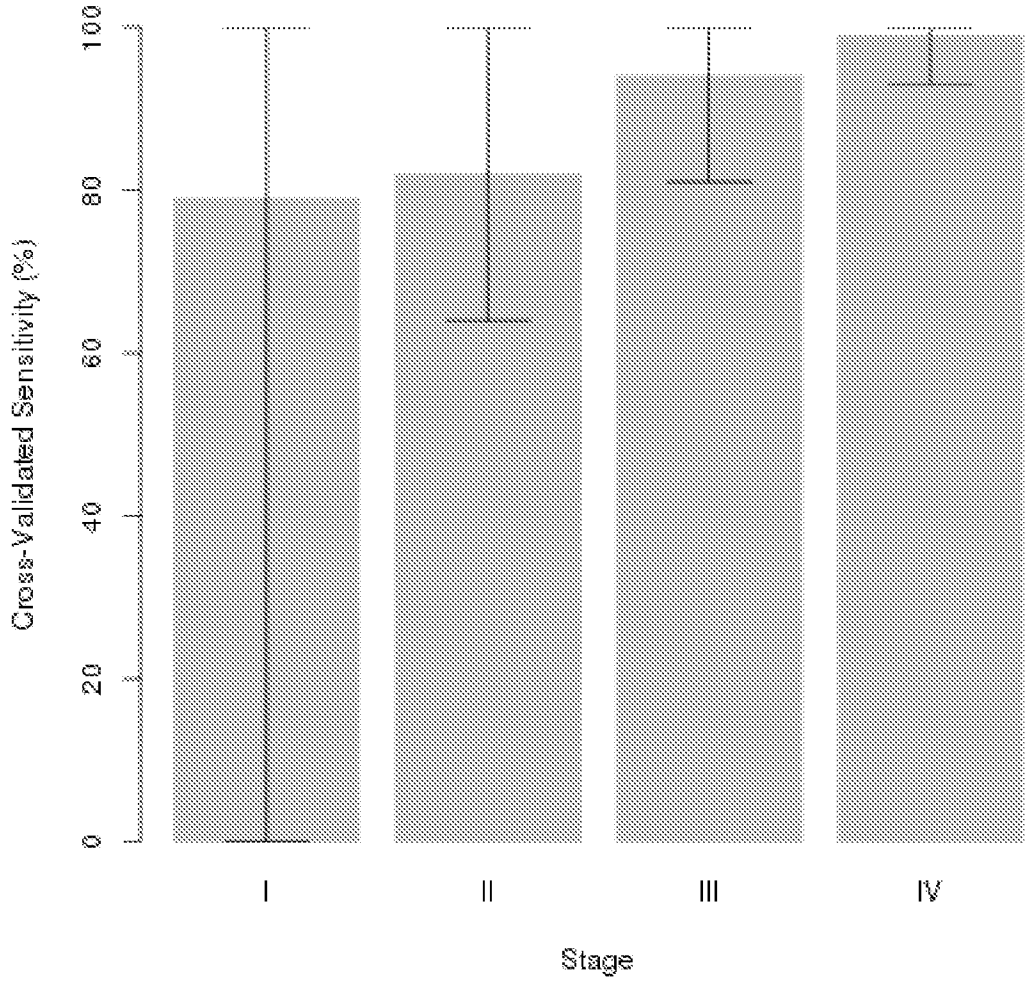
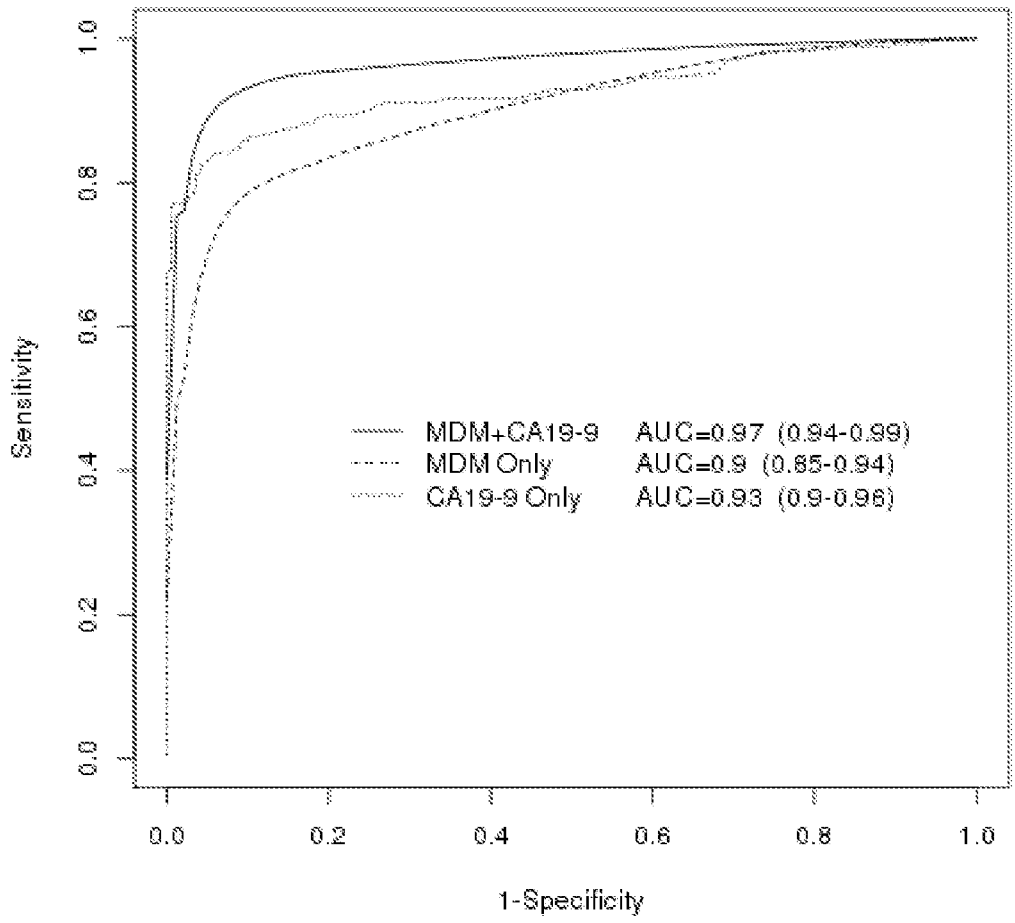

Cross-Validated Sensitivity (%)
0
20
40
60
80
100
I
II
III
IV
Stage

FIG. 3


Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0
MDM+CA19-9 AUC=0.97 (0.94-0.99)
MDM Only AUC=0.9 (0.85-0.94)
CA19-9 Only AUC=0.93 (0.9-0.96)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20/26581

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G01N 33/574; C12N 15/11, 15/117; C12Q 1/68 (2020.01)

CPC - G01N 33/57438, 33/574; A61P 35/00, 1/18; C12Q 1/6886, 1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y ----- A	WO 2017/180886 A1 (MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH) 19 October 2017; page 5, lines 7-28; page 5, line 29 - page 6, line 5; page 6, lines 6-8; page 9, lines 28-33; page 10, lines 21-30; page 12, lines 13-25; page 13, lines 12-16; page 14, line 8; page 15, lines 1-2; page 36, lines 1-4; page 43, lines 27-28; page 51, lines 16-20; claims 7-10	56, 58-59 --- 1-7, 16, 21, 23-27, 29-42, 52, 67-69 --- 8, 17, 22, 28, 43, 53-55, 57
X ----- Y	KISIEL et al. 'Hepatocellular Carcinoma Detection by Plasma Methylated DNA: Discovery, Phase I Pilot, and Phase II Clinical Validation' Hepatology, 2019; Vol . 69, No. 3, pp. 1-21; abstract; page 5, second paragraph; page 7, fifth paragraph; page 9, sixth paragraph. DOI 10.1002/hep.30244	9-15, 18-20 --- 1-7, 16, 21, 23-27, 29-42, 52, 67-69
Y	YI et al. 'Novel Methylation Biomarker Panel for the Early Detection of Pancreatic Cancer' Clin Cancer Res; 2013, pp. OF1-OF12; abstract; page OF10, column 2, first paragraph; Table 3, Table 3 Legend. DOI: 10.1158/1078-0432.CCR-12-3224	67-69
A	US 2008/0254447 A1 (FOEKENS et al.) 16 October 2008; paragraph [0075]	8, 17, 22, 28, 43, 53-55, 57
A	US 2009/0239212 A1 (BEEVER et al.) 24 September 2009; paragraph [0002]	8, 17, 22, 28, 43, 53-55, 57
A	US 2012/0246748 A1 (GUO et al.) 27 September 2012; paragraph [0004]	8, 17, 22, 28, 43, 53-55, 57

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
07 August 2020 (07.08.2020)	31 AUG 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20/26581

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

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

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

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

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

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

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

-***-Please See Supplemental Page-***-

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

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

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

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

1-43, 52-59, 67-69; AK055957 (gene), SEQ ID NO: 1 (forward primer); SEQ ID NO: 2 (reverse primer), and SEQ ID NO: 3 (probe);

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

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

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

Groups I+, Claims 1-43, 52-59, 67-69, AK055957 (gene), SEQ ID NO: 1 (forward primer); SEQ ID NO: 2 (reverse primer), and SEQ ID NO: 3 (probe) are directed toward methods, compositions, and kits for measuring methylation levels to screen for PDAC.

The methods, kits, and compositions will be searched to the extent they encompass AK055957 (first exemplary gene), SEQ ID NO: 1 (first exemplary forward primer); SEQ ID NO: 2 (first exemplary reverse primer), and SEQ ID NO: 3 (first exemplary probe). Applicant is invited to elect additional gene(s), with a first set of specified SEQ ID NOs: for each of a forward primer, reverse primer, and probe, for each gene, and/or additional set(s) of primers(s) and probe(s), with specified SEQ ID NO: for each, such that the sequence of each elected species is fully specified (i.e. no optional or variable bases or substituents), to be searched. Additional gene(s) and associated primer and probe sequence(s) will be searched upon the payment of additional fees. It is believed that claims 1-43 (each in-part), 52-55 (each in-part), 56, 57 (in-part), 58, 59, and 67-69 (each in-part) encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass AK055957 (gene), SEQ ID NO: 1 (forward primer); SEQ ID NO: 2 (reverse primer), and SEQ ID NO: 3 (probe). Applicants must specify the searchable claims that encompass any additionally elected gene(s) and associated sequence(s) and/or set(s) of primer/probe sequence(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be CD1D (gene) SEQ ID NO: 4 (forward primer); SEQ ID NO: 5 (reverse primer), and SEQ ID NO: 6 (probe).

Group II, Claims 44 and 60 are directed toward methods for characterizing a sample and screening for PDAC, comprising comparing two sequences to identify differences in the two sequences.

Group III, Claims 45-51 and 61-66 are directed toward systems for characterizing a sample and screening for PDAC, the system comprising an alert component configured to determine a single value based on a combination of methylation states.

The inventions listed as Groups I+, II and III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Groups I+ include AK055957, not present in either of Groups II or III; the special technical features of Group II include identifying differences between two sequences, not present in any of Groups I+ or III; the special technical features of Group III include a component configured to determine a single value based on a combination of methylation states, not present in any of Groups I+ or II.

Groups I+, II and III share the technical features including: characterizing a sample obtained from a human subject to determine the methylation state of a sample, comparing the sample with a control sample or a reference sample; and screening for PDAC in a sample obtained from a subject to determine the methylation state of a sample; and comparing the sample with a control sample or a reference sample. Groups I+ and II further share the technical features including: a method comprising reacting a nucleic acid comprising a DMR with a bisulfite reagent to produce a bisulfite-reacted nucleic acid.

However, these shared technical features are previously disclosed by WO 2017/180886 A1 to Mayo Foundation for Medical Education and Research (hereinafter 'Mayo').

Mayo discloses characterizing a sample obtained from a human subject to determine the methylation state of a sample (characterizing a sample obtained from a human subject to determine the methylation state of a sample; page 4, lines 27-29), comparing the sample with a control sample or a reference sample (comparing the sample with a control sample or a reference sample; page 4, lines 1-6); and screening for PDAC in a sample obtained from a subject to determine the methylation state of a sample (and screening for PDAC in a sample obtained from a subject to determine the methylation state of a sample; page 4, lines 1-6, lines 11-14, lines 27-29); and comparing the sample with a control sample or a reference sample (comparing the sample with a control sample or a reference sample; page 4, lines 1-6). Groups I+ and II further share the technical features including: a method comprising reacting a nucleic acid comprising a DMR with a bisulfite reagent to produce a bisulfite-reacted nucleic acid (a method comprising reacting a nucleic acid comprising a DMR with a bisulfite reagent to produce a bisulfite-reacted nucleic acid; page 5, lines 19-28).

-***-Continued on Next Supplemental Page-***-

-***-Continued from Previous Supplemental Page-***-:
No technical features are shared between the genes and related primer and probe sequences of Groups I+ and, accordingly, these groups lack unity a priori.

Groups I+ share the technical features including: a method, comprising: measuring a methylation level for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with a reagent that modifies DNA in a methylation-specific manner; amplifying the treated genomic DNA using a set of primers for the selected one or more genes; and determining the methylation level of the one or more genes by polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and target capture; a method of characterizing a sample, comprising: a) measuring an amount of at least one methylated marker gene in DNA from the sample; b) measuring the amount of at least one reference marker in the DNA; and c) calculating a value for the amount of the at least one methylated marker gene measured in the DNA as a percentage of the amount of the reference marker gene measured in the DNA, wherein the value indicates the amount of the at least one methylated marker DNA measured in the sample; a method for characterizing a biological sample comprising: (a) measuring a methylation level of a CpG site for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with bisulfite; amplifying the bisulfite-treated genomic DNA using a set of primers for the selected one or more genes; and determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR; (b) comparing the methylation level to a methylation level of a corresponding set of genes in control samples without PDAC; and (c) determining that the individual has PDAC when the methylation level measured in the one or more genes is higher than the methylation level measured in the respective control samples; a method of screening for PDAC in a sample obtained from a subject, the method comprising: 1) assaying a methylation state of a DNA methylation marker comprising a chromosomal region, and 2) identifying the subject as having PDAC when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have PDAC; and a method for characterizing a sample from a human patient comprising: a) obtaining DNA from a sample of a human patient; b) assaying a methylation state of a DNA methylation marker comprising a chromosomal region; c) comparing the assayed methylation state of the one or more DNA methylation markers with methylation level references for the one or more DNA methylation markers for human patients not having PDAC.

However, these shared technical features are previously disclosed by Mayo, as above.

Mayo discloses a method, comprising: measuring a methylation level for one or more genes in a biological sample of a human individual (a method, comprising: measuring a methylation level for one or more genes in a biological sample of a human individual; page 13, lines 12-16) through treating genomic DNA in the biological sample with a reagent that modifies DNA in a methylation-specific manner (through treating genomic DNA in the biological sample with bisulfite (a reagent that modifies DNA in a methylation-specific manner); page 5, lines 7-28); amplifying the treated genomic DNA using a set of primers for the selected one or more genes (amplifying the treated genomic DNA using a set of primers for the selected one or more genes; page 5, lines 7-28; page 6, lines 6-8); and determining the methylation level of the one or more genes by polymerase chain reaction (determining the methylation level of the one or more genes by polymerase chain reaction; page 5, lines 7-28); a method of characterizing a sample, comprising: a) measuring an amount of at least one methylated marker gene in DNA from the sample (a method of characterizing a sample, comprising: a) measuring an amount of at least one methylated marker gene in DNA from the sample; page 13, lines 12-16); b) measuring the amount of at least one reference marker in the DNA (measuring the amount of at least one reference marker in the DNA; page 12, lines 13-25); and c) calculating a value for the amount of the at least one methylated marker gene measured in the DNA as a percentage of the amount of the reference marker gene measured in the DNA (calculating a value for the amount of the at least one methylated marker gene measured in the DNA as a percentage of the amount of the reference marker gene measured in the DNA; page 5, line 29 - page 6, line 5), wherein the value indicates the amount of the at least one methylated marker DNA measured in the sample (wherein the value indicates the amount of the at least one methylated marker DNA measured in the sample; page 5, line 29 - page 6, line 5; page 13, lines 12-16); a method for characterizing a biological sample comprising: (a) measuring a methylation level of a CpG site for one or more genes in a biological sample of a human individual (a method for characterizing a biological sample comprising: (a) measuring a methylation level of a CpG site for one or more genes in a biological sample of a human individual; page 12, lines 26-28; page 13, lines 12-16) through treating genomic DNA in the biological sample with bisulfite (through treating genomic DNA in the biological sample with bisulfite; page 5, lines 7-28); amplifying the bisulfite-treated genomic DNA using a set of primers for the selected one or more genes (amplifying the bisulfite-treated genomic DNA using a set of primers for the selected one or more genes; page 5, lines 7-28); and determining the methylation level of the CpG site by methylation-specific PCR (and determining the methylation level of the CpG site by methylation-specific PCR; page 5, lines 7-28; page 12, lines 26-28); (b) comparing the methylation level to a methylation level of a corresponding set of genes in control samples without PDAC (comparing the methylation level to a methylation level of a corresponding set of genes in control samples without PDAC; page 4, lines 1-6; page 13, lines 27-32); and (c) determining that the individual has PDAC when the methylation level measured in the one or more genes is higher than the methylation level measured in the respective control samples (determining that the individual has PDAC when the methylation level measured in the one or more genes is higher than the methylation level measured in the respective control samples; page 13, lines 27-32); a method of screening for PDAC in a sample obtained from a subject (a method of screening for PDAC in a sample obtained from a subject; page 8, lines 33-34), the method comprising: 1) assaying a methylation state of a DNA methylation marker comprising a chromosomal region (the method comprising: 1) assaying a methylation state of a DNA methylation marker comprising a chromosomal region; page 8, line 33 - page 9, line 9; page 10, lines 8-11), and 2) identifying the subject as having PDAC when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have PDAC (identifying the subject as having PDAC when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have PDAC; page 8, line 33 - page 9, line 9); and a method for characterizing a sample from a human patient (a method for characterizing a sample from a human patient; page 13, lines 12-16) comprising: a) obtaining DNA from a sample of a human patient (comprising: a) obtaining DNA from a sample of a human patient; page 8, lines 33-34; page 10, lines 21-23); b) assaying a methylation state of a DNA methylation marker comprising a chromosomal region (assaying a methylation state of a DNA methylation marker comprising a chromosomal region; page 10, lines 8-11; page 13, lines 12-16); c) comparing the assayed methylation state of the one or more DNA methylation markers with methylation level references for the one or more DNA methylation markers for human patients not having PDAC (comparing the assayed methylation state of the one or more DNA methylation markers with methylation level references for the one or more DNA methylation markers for human patients not having PDAC; page 4, lines 1-6; page 11, lines 1-5).

Since none of the special technical features of the Groups I+, II and III inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Mayo reference, unity of invention is lacking.