

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2018229294 B2**

(54) Title
Detecting prostate cancer

(51) International Patent Classification(s)
C12N 15/09 (2006.01) **C12Q 1/68** (2018.01)
C04B 20/04 (2006.01)

(21) Application No: **2018229294** (22) Date of Filing: **2018.02.27**

(87) WIPO No: **WO18/160576**

(30) Priority Data

(31)	Number	(32)	Date	(33)	Country
	62/464,800		2017.02.28		US

(43) Publication Date: **2018.09.07**

(44) Accepted Journal Date: **2024.06.13**

(71) Applicant(s)
Exact Sciences Corporation; Mayo Foundation for Medical Education and Research

(72) Inventor(s)
Ahlquist, David A.; Taylor, William R.; Kisiel, John B.; Yab, Tracy C.; Mahoney, Douglas W.; Dukek, Brian A.; Gettman, Matthew T.; Allawi, Hatim T.

(74) Agent / Attorney
Spruson & Ferguson, Level 24, Tower 2 Darling Park, 201 Sussex Street, Sydney, NSW, 2000, AU

(56) Related Art
WO 2013/185779 A2
DEVANEY, J.M., et al. "Identification of novel DNA-methylated genes that correlate with human prostate cancer and high-grade prostatic intraepithelial neoplasia", PROSTATE CANCER PROSTATIC DIS, 2013, vol. 16, no. 4, pp 292-300



(51) International Patent Classification:

C12N 15/09 (2006.01) C04B 20/04 (2006.01)

C12Q 1/68 (2018.01)

(21) International Application Number:

PCT/US2018/019982

(22) International Filing Date:

27 February 2018 (27.02.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/464,800 28 February 2017 (28.02.2017) 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: AHLQUIST, David A.; c/o Mayo Foundation for Medical Education & Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). TAYLOR, William R.; c/o Mayo Foundation for Medical Education & Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). KISIEL, John B.; c/o Mayo Foundation for Medical Education & Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). YAB, Tracy C.; c/o Mayo Foundation for Medical Education & Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). MAHONEY, Douglas W.; c/o Mayo Foundation for Medical Education & Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). DUKEK, Brian A.; c/o Mayo Foundation for Medical Education & Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). GETTMAN, Matthew T.; c/o Mayo Foundation for Medical Education & 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

(54) Title: DETECTING PROSTATE CANCER

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



WO 2018/160576 A1

DETECTING PROSTATE CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims priority to U.S. Provisional Patent Application No. 5 62/464,800, filed February 28, 2017, the contents of which are incorporated by reference in its entirety.

FIELD OF INVENTION

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

BACKGROUND

Prostate cancer (PCa) is the second most commonly diagnosed cancer in men 15 representing 903,000 new cases and 258,000 deaths worldwide in 2008. While PCA is common, the disease is also heterogeneous in clinical behavior. It is estimated that approximately 1 in 6 American men will be diagnosed with PCA, but the PCA death rate in American men is only 2.8% (1 in 36 men) (see, e.g., Strand SH, et al., Int J Mol Sci 2014;15:16544-16576).

20 PCA survival depends on many factors. An earlier diagnosis of less advanced disease provides most men the best chance of curative treatment. Indeed, an earlier PCA diagnosis has been facilitated with the use of prostate specific antigen (PSA) testing. In the PSA testing era, a stage and grade migration of PCA has occurred leading to identification of disease more amenable to definitive treatment. While PSA has been beneficial in the diagnosis and 25 management of PCA, screening with PSA has also been viewed as controversial. Screening with PSA may lead to the diagnosis of more indolent, low risk tumors and subsequent overtreatment may subject men to unnecessary quality-of-life harms (erectile dysfunction, incontinence). Accordingly, new biomarkers are needed to assist with the diagnosis of PCA. New tests are also needed to provide men with improved prognostic information regarding 30 their cancer.

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 promoter 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 microarray based interrogation of CpG methylation is a reasonable, high-throughput approach, this strategy is biased towards known regions of interest, mainly established tumor suppressor promoters. 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 prostate cancer screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of prostate cancer.

Indeed, as described in Examples I-VIII, experiments conducted during the course for identifying embodiments for the present invention identified a novel set of 73 differentially methylated regions (DMRs) for discriminating cancer of the prostate derived DNA from non-

neoplastic control DNA. In addition, 10 novel DMRs were identified which are methylated in prostate epithelia (cancer and normal) but not in normal leukocyte DNA samples. Both sets of regions were identified from next generation sequencing studies on CpG enriched bisulfite converted tumor and normal DNA. Tumor samples included less aggressive Gleason 6 and more aggressive Gleason 7+ patterns. DMRs were selected using proprietary filters and analysis pipelines and validated in independent tissue sample sets using novel methylation-specific PCR (MSP) assays. These 73 biomarker assays demonstrated superior detection in tissues and have a broad spectrum of clinical specificity - some for cancers across many different organ sites, others specific to prostate cancer only.

Such experiments list and describe 120 novel DNA methylation markers (Table 1) distinguishing prostate cancer tissue from benign prostate tissue. From these 120 novel DNA methylation markers, further experiments identified 73 markers capable of distinguishing aggressive prostate cancer tissue (e.g., Gleason Score 7+) from benign prostate tissue. More specifically, markers and/or panels of markers were identified (e.g., a chromosomal region having an annotation selected from ACOXL, AKR1B1_3644, ANXA2, CHST11_2206, FLJ45983, GAS6, GRASP, HAPLN3, HCG4P6, HES5_0822, ITPRIPL1, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.193, MAX.chr3.72788028-72788112, RAI1_7469, RASSF2, SERPINB9_3389, SLC4A11, and TPM4_8047) capable of distinguishing prostate cancer tissue from benign prostate tissue (see, Examples I-VI).

Additional experiments conducted during the course of developing embodiments for the present invention were directed toward identifying markers capable of distinguishing prostate cancer tissue from benign prostate tissue (e.g., a chromosomal region having an annotation selected from SERPINB9_3479, FLOT1_1665, HCG4P6_4618, CHST11_2206, MAX.chr12.485, GRASP_0932, GAS6_6425, MAX.chr3.193, MAX.chr2.971_3164, MAX.chr3.727_8028, HES5_0840, TPM4_8037, SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487) capable of distinguishing prostate cancer tissue from benign prostate tissue (see, Example VIII; Table 11).

Additional experiments conducted during the course of developing embodiments for the present invention were directed toward identifying markers capable of distinguishing aggressive prostate cancer tissue (e.g., Gleason Score 7+) from less aggressive prostate cancer tissue (e.g., Gleason Score 6) (e.g., a chromosomal region having an annotation

selected from SERPINB9_3479, GRASP_0932, SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487) capable of distinguishing prostate cancer tissue from benign prostate tissue (see, Example VIII; Table 11).

5 Additional experiments conducted during the course of developing embodiments for the present invention were directed toward identifying markers capable of detecting the presence or absence of prostate cancer within blood samples (e.g., blood plasma samples). Indeed, markers and/or panels of markers were identified (e.g., a chromosomal region having an annotation selected from max.chr3.193, HES5, SLCO3A1, and TPM4_8047) capable of
10 detecting the presence or absence of prostate cancer tissue within blood plasma samples (see, Examples I-VI).

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, or 8 markers) with high discrimination for prostate cancer overall. Experiments applied a selection filter to candidate markers to identify
15 markers that provide a high signal to noise ratio and a low background level to provide high specificity for purposes of prostate cancer 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., prostate tissue, plasma sample). These markers comprise one or more differentially
20 methylated regions (DMR) as discussed herein, e.g., as provided in Tables 1 and 3. 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
25 (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
30 some embodiments, analyzing changes in methylation patterns involves a PCR-based process that involves digestion of genomic DNA with methylation-sensitive 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
5 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
10 “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,
15 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
20 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
25 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
30 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., $(\text{number of methylated DNAs}) / (\text{the number of methylated DNAs} + \text{number of unmethylated DNAs}) \times 100$.

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 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 1-140 as provided in Tables 1 and 13); comparing methylation states (e.g., of one or more DMR, e.g., DMR 1-140 as provided in Tables 1 and 13); generating standard curves; determining a Ct value; calculating a fraction, frequency, or percentage of methylation (e.g., of one or more DMR, e.g., (e.g., of one or more DMR, e.g., DMR 1-140 as provided in Tables 1 and 13); 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 Tables 1 and 3). 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 Tables 1 and 3). 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 prostate cancer 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., prostate tissue) (e.g., plasma sample) and identifying the subject as having prostate cancer when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have prostate cancer, wherein the marker comprises a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1–140 as provided in Tables 1 and 13.

In some embodiments wherein the sample obtained from the subject is prostate tissue, the marker is selected from ACOXL, AKR1B1_3644, ANXA2, CHST11_2206, FLJ45983, GAS6, GRASP, HAPLN3, HCG4P6, HES5_0822, ITPRIPL1, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.193, MAX.chr3.72788028-72788112, RAI1_7469, RASSF2, SERPINB9_3389, SLC4A11, and TPM4_8047.

In some embodiments wherein the sample obtained from the subject is prostate tissue, the marker is selected from SERPINB9_3479, FLOT1_1665, HCG4P6_4618, CHST11_2206, MAX.chr12.485, GRASP_0932, GAS6_6425, MAX.chr3.193, MAX.chr2.971_3164, MAX.chr3.727_8028, HES5_0840, TPM4_8037, SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487.

In some embodiments wherein the sample obtained from the subject is blood plasma, the marker is selected from max.chr3.193, HES5, SLCO3A1, and TPM4_8047.

The technology is related to identifying and discriminating prostate cancer. Some embodiments provide methods comprising assaying a plurality of markers, e.g., comprising assaying 2 to 11 to 100 or 140 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
 5 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.
 10 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., prostate tissue sample), a blood sample (e.g., plasma, serum, whole blood), an excretion, or a urine sample.

Furthermore, the technology is not limited in the method used to determine
 15 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
 20 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: 1–146 and/or SEQ ID NO: 147-234. In some embodiments are provided an oligonucleotide
 25 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, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ACOXL, AKR1B1_3644, ANXA2, CHST11_2206, FLJ45983, GAS6, GRASP, HAPLN3, HCG4P6,
 30 HES5_0822, ITPRIPL1, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.193, MAX.chr3.72788028-72788112, RAI1_7469, RASSF2, SERPINB9_3389, SLC4A11, and TPM4_8047, and that comprises the marker (see, Examples I-VI).

In some embodiments the marker comprises a chromosomal region having an annotation that is SERPINB9_3479, FLOT1_1665, HCG4P6_4618, CHST11_2206, MAX.chr12.485, GRASP_0932, GAS6_6425, MAX.chr3.193, MAX.chr2.971_3164, MAX.chr3.727_8028, HES5_0840, TPM4_8037, SLCO3A1_6187, ITPRIPL1_1244,
 5 AKR1B1_3644, RASGRF2_6325, ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487, and that comprises the marker (see, Examples VIII).

In some embodiments wherein the obtained sample is a plasma sample, the marker comprises a chromosomal region having an annotation that is max.chr3.193, HES5,
 10 SLCO3A1, and TPM4_8047, and that comprises the marker.

Kit embodiments are provided, e.g., a kit comprising a bisulfite reagent; and a control nucleic acid comprising a sequence from a DMR selected from a group consisting of DMR 1–140 (from Tables 1 or 13) and having a methylation state associated with a subject who does not have prostate cancer. In some embodiments, kits comprise a bisulfite reagent and an
 15 oligonucleotide as described herein. In some embodiments, kits comprise a bisulfite reagent; and a control nucleic acid comprising a sequence from a DMR selected from a group consisting of DMR 1–140 (from Tables 1 or 13) and having a methylation state associated with a subject who has prostate cancer. Some kit embodiments comprise a sample collector for obtaining a sample from a subject (e.g., a stool sample; prostate tissue sample; plasma
 20 sample); reagents for isolating a nucleic acid from the sample; 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 bisulfite reagent. Some embodiments provide a composition comprising a nucleic
 25 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 prostate cancer
 30 in a sample obtained from a subject (e.g., prostate tissue sample; plasma 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 1–140 (from Tables 1 or 13); 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 prostate cancer; 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 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 prostate cancer to identify differences in the two sequences; and identifying the subject as having prostate cancer when a difference is present.

Systems for screening for prostate cancer in a sample obtained from a subject are provided by the technology. Exemplary embodiments of systems include, e.g., a system for screening for prostate cancer in a sample obtained from a subject (e.g., prostate 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 prostate-cancer-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 Tables 1 or 3) 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 prostate cancer. 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 prostate cancer. Related system embodiments comprise a set of nucleic acids as described and
5 a database of nucleic acid sequences associated with the set of nucleic acids. Some embodiments further comprise a bisulfite reagent. And, some embodiments further comprise a nucleic acid sequencer.

In certain embodiments, methods for characterizing a sample (e.g., prostate tissue sample; plasma sample; stool sample) from a human patient are provided. For example, in
10 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 1–140 from Tables 1 or 13; 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
15 markers for human patients not having prostate cancer.

Such methods are not limited to a particular type of sample from a human patient. In some embodiments, the sample is a prostate tissue sample. In some embodiments, the sample is a plasma sample. In some embodiments, the sample is a stool sample, a tissue sample, a prostate tissue sample, a blood sample, or a urine sample.

20 In some embodiments, such methods comprise assaying a plurality of DNA methylation markers. In some embodiments, such methods comprise assaying 2 to 11 DNA methylation markers. In some embodiments, such methods comprise assaying 12 to 140 DNA methylation markers. In some embodiments, such methods comprise assaying the methylation state of the one or more DNA methylation markers in the sample comprises
25 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.

30 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,
5 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: 1–146 and/or SEQ ID NO: 147–234.

In some embodiments, a chromosomal region having an annotation selected from the
10 group consisting of ACOXL, AKR1B1_3644, ANXA2, CHST11_2206, FLJ45983, GAS6, GRASP, HAPLN3, HCG4P6, HES5_0822, ITPRIPL1, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.193, MAX.chr3.72788028-72788112, RAI1_7469, RASSF2, SERPINB9_3389, SLC4A11, and TPM4_8047 comprises the DNA methylation marker. In some embodiments, the DMR is from Table 3.

In some embodiments, a chromosomal region having an annotation selected from the
15 group consisting of SERPINB9_3479, FLOT1_1665, HCG4P6_4618, CHST11_2206, MAX.chr12.485, GRASP_0932, GAS6_6425, MAX.chr3.193, MAX.chr2.971_3164, MAX.chr3.727_8028, HES5_0840, TPM4_8037, SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113,
20 CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487 comprises the DNA methylation marker.

In some embodiments wherein the obtained sample is a plasma sample, the marker comprises a chromosomal region having an annotation that is max.chr3.193, HES5, SLCO3A1, and TPM4_8047, and that comprises the marker.

In some embodiments, such methods comprise determining the methylation state of
25 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 Tables 1 or 3.

In certain embodiments, the technology provides methods for characterizing a sample
30 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 1–140 from Tables 1 and 13; 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 a prostate cancer; 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., prostate tissue sample; plasma sample; stool sample), 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 prostate cancer 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., prostate 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 prostate cancer-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 prostate tissue sample, a blood 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 a prostate cancer.

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

DETAILED DESCRIPTION

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

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
10 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.

All literature and similar materials cited in this application, including but not limited to, patents, patent applications, articles, books, treatises, and internet web pages are expressly incorporated by reference in their entirety for any purpose. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood
15 by one of ordinary skill in the art to which the various embodiments described herein belongs. When definitions of terms in incorporated references appear to differ from the definitions provided in the present teachings, the definition provided in the present teachings shall control.

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
25 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
30 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.”

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”
5 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
10 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
15 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,
20 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
25 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
30 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.

Amplification of nucleic acids generally 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, 10 to 100 copies of a polynucleotide molecule, which may or may not be exactly the same), 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. (1997) *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
5 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 DNA without
10 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
15 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
20 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
25 terms of concentration) in the mixture, they are said to be “PCR amplified” and are “PCR products” or “amplicons.”

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.
30 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 2009/0253142, 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; 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., Barnay 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.”

5 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
10 be present as background in a test sample.

The term “primer” refers to an oligonucleotide, whether occurring naturally as in a purified 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 strand is induced, (e.g., in the presence of
15 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
20 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,
25 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
30 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.

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 or absence of one or more methylated nucleotide bases in the nucleic acid molecule. For example, a nucleic acid molecule containing a methylated cytosine is considered methylated (e.g., the methylation state of the nucleic acid molecule is methylated). A nucleic acid molecule that does not contain any methylated nucleotides is considered unmethylated.

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
5 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
10 “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
15 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.

20 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
25 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
30 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 $R = (A \times B) / (C \times D)$, where R is the ratio of observed CpG frequency over expected frequency, A is the number of CpG dinucleotides in an analyzed sequence, B is the total number of nucleotides in the analyzed sequence, C is the total number of C nucleotides in the analyzed sequence, and D is the total number of G nucleotides in the analyzed sequence. Methylation state is typically determined in CpG islands, e.g., at promoter regions. It will be appreciated though that other sequences in the human genome are prone to DNA methylation such as CpA and CpT (see Ramsahoye (2000) *Proc. Natl. Acad. Sci. USA* **97**: 5237–5242; Salmon and Kaye (1970) *Biochim. Biophys. Acta.* **204**: 340–351; Grafstrom (1985) *Nucleic Acids Res.* **13**: 2827–2842; Nyce (1986) *Nucleic Acids Res.* **14**: 4353–4367; Woodcock (1987) *Biochem. Biophys. Res. Commun.* **145**: 888-894).

As used herein, a 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 a change in the nucleic acid molecule's nucleotide sequence can result in a nucleic acid molecule in which each methylated nucleotide is modified to a different nucleotide. Such a change in the nucleic acid nucleotide sequence can result in a nucleic acid molecule in which each unmethylated nucleotide is modified to a different nucleotide. Such a change in the nucleic acid nucleotide sequence can result in a nucleic acid molecule in which each of a selected nucleotide which is unmethylated (e.g., each unmethylated cytosine) is modified to a different nucleotide. Use of such a reagent to change the nucleic acid nucleotide sequence can result in a nucleic acid molecule in which each nucleotide that is a methylated nucleotide (e.g., each methylated cytosine) is modified to a different nucleotide. As used herein, use of a reagent that modifies a selected nucleotide refers to a reagent that modifies 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), such that the reagent modifies the one nucleotide without modifying the other three nucleotides. In one exemplary embodiment, such a reagent modifies an unmethylated selected nucleotide to produce a different nucleotide. In another exemplary embodiment, such a reagent can deaminate unmethylated cytosine nucleotides. An exemplary reagent is bisulfite.

As used herein, the term "bisulfite reagent" refers to a reagent comprising in some embodiments bisulfite, disulfite, hydrogen sulfite, or combinations thereof to distinguish between methylated and unmethylated cytidines, e.g., in CpG dinucleotide sequences.

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.

5 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.

10 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.

15 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.

20

The terms “methylation-specific restriction enzyme” or “methylation-sensitive restriction enzyme” refers to an 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 hemimethylated, the cut will not take place or will take place with a significantly reduced efficiency if the recognition site is methylated. In the case of a restriction enzyme that specifically cuts if the recognition site is methylated, 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

25

30

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-pairing 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 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 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).

As used herein, the term “neoplasm” refers to “an abnormal mass of tissue, the growth of which exceeds and is uncoordinated with that of the normal tissues” See, e.g., Willis RA, “The Spread of Tumors in the Human Body”, London, Butterworth & Co, 1952.

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.

5 The term “marker”, as used herein, refers to a substance (e.g., a nucleic acid or a region of a nucleic acid) that is able to diagnose a cancer by distinguishing cancerous cells from normal cells, e.g., based its methylation state.

 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
10 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;
15 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
20 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
25 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
30 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
5 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.,
10 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 is meant to include a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained
15 from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. In some embodiments, the sample is a plasma sample. In some embodiments, the sample is a prostate tissue sample. In some embodiments, the sample is a stool sample. 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
20 the present invention.

As used herein, a “remote sample” as used in some contexts relates to a sample indirectly collected from a site that is not the cell, tissue, or organ source of the sample. For instance, when sample material originating from the pancreas is assessed in a stool sample (e.g., not from a sample taken directly from a prostate), the sample is a remote sample.

25 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.

As used herein, the term “kit” refers to any delivery system for delivering materials.
30 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.

Embodiments of the technology

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

Indeed, as described in Examples I-VI, experiments conducted during the course for identifying embodiments for the present invention identified a novel set of 73 differentially methylated regions (DMRs) for discriminating cancer of the prostate derived DNA from non-neoplastic control DNA. In addition, 10 novel DMRs were identified which are methylated in prostate epithelia (cancer and normal) but not in normal leukocyte DNA samples. Both sets of regions were identified from next generation sequencing studies on CpG enriched bisulfite converted tumor and normal DNA. Tumor samples included less aggressive Gleason 6 and more aggressive Gleason 7+ patterns. DMRs were selected using proprietary filters and analysis pipelines and validated in independent tissue sample sets using novel methylation-specific PCR (MSP) assays. These 73 biomarker assays demonstrated superior detection in tissues and have a broad spectrum of clinical specificity - some for cancers across many different organ sites, others specific to prostate cancer only.

Such experiments list and describe 120 novel DNA methylation markers (Table 1) distinguishing prostate cancer tissue from benign prostate tissue. From these 120 novel DNA methylation markers, further experiments identified 73 markers capable of distinguishing

aggressive prostate cancer tissue (e.g., Gleason Score 7+) from benign prostate tissue. More specifically, markers and/or panels of markers were identified (e.g., a chromosomal region having an annotation selected from ACOXL, AKR1B1_3644, ANXA2, CHST11_2206, FLJ45983, GAS6, GRASP, HAPLN3, HCG4P6, HES5_0822, ITPRIPL1, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.193, MAX.chr3.72788028-72788112, RAI1_7469, RASSF2, SERPINB9_3389, SLC4A11, and TPM4_8047) capable of distinguishing prostate cancer tissue from benign prostate tissue (see, Examples I-VI).

Additional experiments conducted during the course of developing embodiments for the present invention were directed toward identifying markers capable of distinguishing prostate cancer tissue from benign prostate tissue (e.g., a chromosomal region having an annotation selected from SERPINB9_3479, FLOT1_1665, HCG4P6_4618, CHST11_2206, MAX.chr12.485, GRASP_0932, GAS6_6425, MAX.chr3.193, MAX.chr2.971_3164, MAX.chr3.727_8028, HES5_0840, TPM4_8037, SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487) capable of distinguishing prostate cancer tissue from benign prostate tissue (see, Example VIII; Table 11).

Additional experiments conducted during the course of developing embodiments for the present invention were directed toward identifying markers capable of distinguishing aggressive prostate cancer tissue (e.g., Gleason Score 7+) from less aggressive prostate cancer tissue (e.g., Gleason Score 6) (e.g., a chromosomal region having an annotation selected from SERPINB9_3479, GRASP_0932, SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487) capable of distinguishing prostate cancer tissue from benign prostate tissue (see, Example VIII; Table 11).

Additional experiments conducted during the course of developing embodiments for the present invention were directed toward identifying markers capable of detecting the presence or absence of prostate cancer within blood samples (e.g., blood plasma samples). Indeed, markers and/or panels of markers were identified (e.g., a chromosomal region having an annotation selected from max.chr3.193, HES5, SLCO3A1, and TPM4_8047) capable of detecting the presence or absence of prostate cancer tissue within blood plasma samples (see, Examples I-VI).

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.

5 In particular aspects, the present technology provides compositions and methods for identifying, determining, and/or classifying a cancer such as prostate cancer. 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, prostate tissue sample, plasma sample), wherein a change in the methylation state of the marker is indicative of the presence, class, or site of a prostate cancer. Particular embodiments relate to markers comprising a
10 differentially methylated region (DMR, e.g., DMR 1–140, see Tables 1 and 13) that are used for diagnosis (e.g., screening) of prostate cancer.

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 1–140) provided herein and listed in Tables 1 or 3 is analyzed, the technology also provides panels of
15 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 prostate cancer.

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.

20 In some embodiments, the present technology provides for the use of the bisulfite technique 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 1–140, see Tables 1 and 13). Genomic CpG dinucleotides can be methylated or unmethylated (alternatively known as up- and down-methylated respectively). However the
25 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,
30 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 prostate cancer.

Combinations of markers

In some embodiments, the technology relates to assessing the methylation state of combinations of markers comprising a DMR from Table 1, Table 3 or Table 13 (e.g., DMR Nos. 1-140). 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., prostate cancer).

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

The most 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 forms 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) or methylation-specific PCR (MSP) as is disclosed, e.g., in U.S. Patent No. 5,786,146.

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.

5 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 (Gonzalzo & 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
10 (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.; Zeschneigk 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;
15 WO 9746705; WO 9515373).

Various methylation assay procedures are known in the art and 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,
20 sequencing of bisulfite-treated nucleic acid, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction 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
25 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
30 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.

Preferably, 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 (Gonzalzo & 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, DMR, 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, DMR, 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, overlap any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by
5 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
10 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
15 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
20 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, DMR, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, etc.); TaqMan® or Lightcycler® probes; optimized PCR buffers
25 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
30 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, overlap 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, DMR, 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 (Gonzalzo & 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, DMR, 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.

5 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
10 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.
15 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,
20 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
25 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
30 amplicon. The presence of the specific invasive oligonucleotide at the target binding site causes cleavage to release the flap sequence by cutting between the detection probe and the flap sequence. The flap sequence is complementary to a nonhairpin 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; U.S. Pat. Appl. Ser. Nos. 12/946,737, 12/946,745, 12/946,752, and 61/548,639.

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, which is incorporated by reference in its entirety). It is preferred that the bisulfite treatment is conducted in the presence of denaturing solvents such as but not limited to n-alkylenglycol 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 derivatives thereof, e.g., Gallic acid (see: PCT/EP2004/011715, which is incorporated by reference in its entirety). The bisulfite conversion is preferably carried out at a reaction temperature between 30°C and 70°C, whereby the temperature is increased to over 85°C for short times during the reaction (see: PCT/EP2004/011715, which is incorporated by reference in its entirety). The bisulfite treated DNA is preferably purified prior to the quantification. This may be conducted by any means known in the art, such as but not limited to ultrafiltration, e.g., by means of Microcon™ columns (manufactured by Millipore™). The purification is carried out according to a modified manufacturer's protocol (see, e.g., PCT/EP2004/011715, which is incorporated by reference in its entirety).

In some embodiments, fragments of the treated DNA are amplified using sets of primer oligonucleotides according to the present invention (e.g., see Tables 3 and/or 5) 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 1–140 as provided in Tables 1 and 13) 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 detectability 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.

Methods

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

- 1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from a body fluids such as a stool sample or prostate tissue or plasma 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 comprising a DMR (e.g., DMR 1–140, e.g., as provided in Tables 1 and 13) and

- 2) detecting prostate cancer (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 the technology, methods are provided that comprise the
 5 following steps:

- 1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from a body fluids such as a stool sample or prostate tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG
 10 dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ACOXL, AKR1B1_3644, ANXA2, CHST11_2206, FLJ45983, GAS6, GRASP, HAPLN3, HCG4P6, HES5_0822, ITPRIPL1, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.193, MAX.chr3.72788028-72788112,
 15 RAI1_7469, RASSF2, SERPINB9_3389, SLC4A11, and TPM4_8047, and
- 2) detecting prostate cancer (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 the technology, methods are provided that comprise the
 20 following steps:

- 1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from a body fluids such as a stool sample or prostate tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG
 25 dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of SERPINB9_3479, FLOT1_1665, HCG4P6_4618, CHST11_2206, MAX.chr12.485, GRASP_0932, GAS6_6425, MAX.chr3.193, MAX.chr2.971_3164, MAX.chr3.727_8028, HES5_0840, TPM4_8037, SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325,
 30 ZNF655_6075, PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and WNT3A_5487, and
- 2) detecting prostate cancer (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 the technology, methods are provided that comprise the following steps:

- 5 1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from a plasma 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 max.chr3.193, HES5, SLCO3A1, and TPM4_8047, and
- 10 2) detecting prostate cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

Preferably, the sensitivity 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 15 100%, 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
20 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
25 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, prostate tissue, colonic effluent, urine, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood, and combinations thereof.

The technology is not limited in the methods used to prepare the samples and provide
30 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 1–140, e.g., as provided by Tables 1 and 13).

5 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.

10 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.

15 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 1–140, e.g., as provided in Tables 1 and 13). 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.

20 Aberrant methylation, more specifically hypermethylation of a marker comprising a DMR (e.g., DMR 1–140, e.g., as provided by Tables 1 and 13) is associated with a prostate cancer.

 The technology relates to the analysis of any sample associated with a prostate cancer. 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
25 embodiments, the sample comprises blood, serum, plasma, gastric secretions, pancreatic juice, a gastrointestinal biopsy sample, microdissected cells from a prostate biopsy, and/or prostate cells recovered from stool. In some embodiments, the subject is human. The sample may include cells, secretions, or tissues from the prostate, liver, bile ducts, pancreas, stomach, colon, rectum, esophagus, small intestine, appendix, duodenum, polyps, gall bladder, anus,
30 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.

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
5 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
10 patient with prostate cancer, with early stage prostate cancer, or who may develop prostate cancer), 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.
15 Preferably, said use is in a method of clinical screening, a method of prognosis assessment, a method of monitoring the results of therapy, a method to identify patients most likely to respond to a particular therapeutic treatment, a method of imaging a patient or subject, and a method for drug screening and development.

In some embodiments of the technology, a method for diagnosing a prostate cancer in
20 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
25 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
30 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 prostate cancer, and/or monitor the efficacy of appropriate therapies directed against the cancer. In such an embodiment for example, one might expect to see a change in the methylation state of one or more biomarkers (e.g., DMR) disclosed herein (and potentially one or more additional biomarker(s), if monitored) in a biological sample 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 prostate cancer risk, prognosis, determining treatment efficacy, and/or progression of the cancer in the subject.

In preferred embodiments, the methods and compositions of the invention are for treatment or diagnosis of disease 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
5 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.

10 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.

15 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
20 will understand that the term “prognosis” refers to an increased probability that a certain course or outcome will occur; that is, that a course or outcome is more likely to occur in a subject exhibiting a given condition, when compared to those individuals not exhibiting the condition. For example, in individuals not exhibiting the condition (e.g., having a normal methylation state of one or more DMR), the chance of a given outcome (e.g., suffering from a
25 prostate cancer) may be very low.

In some embodiments, a statistical analysis associates a prognostic indicator with a predisposition to an adverse outcome. For example, in some embodiments, a methylation state different from that in a normal control sample obtained from a patient who does not have a cancer can signal that a subject is more likely to suffer from a cancer than subjects
30 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%,
5 99%, 99.5%, 99.9% and 99.99%, while exemplary *p* values are 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, and 0.0001.

In other embodiments, a threshold degree of change in the methylation state of a prognostic or diagnostic biomarker disclosed herein (e.g., a DMR) can be established, and the degree of change in the methylation state of the biomarker in a biological sample is simply
10 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
15 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.,
25 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 a prostate cancer. In certain embodiments of the method, a subject is identified as having metaplasia upon identifying an
30 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
5 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
10 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
15 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 a prostate cancer if, when compared to a control methylation state, there is a measurable difference in the methylation
20 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 prostate cancer, 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 a prostate
25 cancer 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 an endoscopy, until such time as a future screening, for example, a screening conducted in accordance with the present technology, indicates that a risk of prostate cancer has appeared in those subjects.

30 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, a prostate cancer 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
5 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
10 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
15 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
20 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
25 further includes a system for diagnosing a prostate 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 prostate cancer or diagnose a prostate 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 DMR as provided in Tables 1 and 3.

EXAMPLES

Example I.

This example provides the materials and methods for Examples II, III, IV, V and VI.

Examples II, III, IV, V and VI demonstrate that methylated DNA markers can discriminate prostate tissue (e.g., cancerous prostate tissue and/or non-cancerous prostate tissue) from non-prostate tissue (e.g., leukocyte cells), that methylated DNA markers can discriminate cancerous prostate tissue from non-cancerous prostate tissue, that methylated DNA markers can discriminate highly aggressive cancerous prostate tissue (e.g., Gleason score at or above 7.0 (e.g., 7, 8, 9, 10) from less-aggressive cancerous prostate tissue (e.g., Gleason score below 7 (e.g., 6), and that methylated DNA markers can detect PCa in blood samples.

These experiments comprised five phases. First, DNA methylation marker discovery was performed using Reduced Representation Bisulfite Sequencing (RRBS) (see, e.g., Gu H, et al., Nat Methods 2010;7:133-6) on DNA extracted from prostate cancer (PCa) tissues (both Gleason score 6 and 7+), from normal prostate and from buffy coat samples from healthy volunteers. Second, discriminant differentially methylated regions (DMRs) were identified by strict filtration criteria and the sequences used to develop real-time methylation specific PCR assays (qMSP). These assays were then applied to the original sample set to ensure reproducibility of results (technical validation). Third, best performing candidate markers were selected for qMSP biological validation on DNA extracted from independent archival case and control tissues. Fourth, candidate marker sequences were compared in-silico across a pan-cancer RRBS sequencing data set to gauge the degree of site specific methylation for each marker. Fifth, a set of high performing PCa markers chosen for testing in blinded independent plasma samples to assess PCa detection in a clinical medium.

Study subjects and samples

The study was approved by the Mayo Clinic Institutional Review Board (Rochester, MN). Fresh frozen (FF) tissues, plasma, and buffy coat samples were provided by IRB-approved patient biobanks. Tumor tissue sections were re-reviewed by an expert GI pathologist to confirm diagnosis and estimate neoplastic cellularity. Sections were then macro-dissected. Genomic DNA was purified using the QiaAmp Mini kit (Qiagen, Valencia CA) and subsequently re-purified with the AMPure XP kit (Beckman Coulter, Brea CA).

Preparation of Reduced Representation Bisulfite Sequencing Libraries

150 ng of each sample's DNA was diluted in 26 ul of Te buffer (5.77 ng/ul). This was digested overnight by 1 ul (20 units) of MspI in 1X final concentration CutSmart buffer (New England Biolabs). The 3' overhangs were end repaired and A tailed with a mixture of 0.6 ul 100 mM dATP, 0.06 ul 100 mM dCTP, and 0.06 ul dGTP with 2 ul (10 units) Klenow DNA polymerase (New England Biolabs). Product was incubated for 20 minutes at 30 degrees, 20 minutes at 37 degrees, and held at 4 degrees. Following end repair, product was purified by 2X Agencourt Ampure XP beads (Beckman Coulter), washed twice with 70% EtOH, and eluted in 20 ul of water. Illumina adaptors were ligated to the product using 1 ul T4 ligase (400 units) in 1X T4 ligase buffer incubated at 16 degrees overnight. The product was treated at 65 degrees for 20 minutes to heat inactivate the enzyme. Following ligation, product was purified by 2X Agencourt Ampure XP beads (Beckman Coulter), washed twice with 70% EtOH, and eluted in 47 ul of water. 45 ul of product was bisulfite converted by the EZ-96 DNA methylation kit (Zymo Research) as described in their protocol. Converted product was purified by 2X Agencourt Ampure XP beads (Beckman Coulter), washed twice with 70% EtOH, and eluted in 22 ul of water. Illumina indexes were added via PCR using 16 ul of bisulfite converted product, 1 ul (2.5 units) of PfuTurbo Cx hotstart DNA polymerase, 0.5 ul dNTPs (25 mM each), 6 ul Illumina indexes (2.5 uM each), and 1X PfuTurbo Cx hotstart DNA polymerase buffer in a 50 ul total volume. Product was dual size selected by using Ampure XP at 0.7X and collecting the supernatant while followed by Ampure XP at 1.2X and reserving what bound to the beads. Final product was eluted in 40 ul. DNA mass yield was determined by Pico Green (Molecular Probes) and measured on a Tecan fluorometer. DNA size was determined by High Sensitivity DNA chip on the Agilent 2100 (Agilent). Molarity of products were calculated and samples were pooled equimolarly at four samples/pool to 10 nM.

Massively Parallel Sequencing and Bioinformatics

Samples were loaded onto flow cells according to a randomized lane assignment. Sequencing was performed by the Next Generation Sequencing Core at the Mayo Clinic Medical Genome Facility on the Illumina HiSeq 2000. Reads were unidirectional for 101 cycles. Each flow cell lane generated 100-120 million reads, sufficient for a median coverage of 30-50 fold sequencing depth for aligned sequences. Standard Illumina pipeline software called bases and generated reads in the fastq format. SAAP-RRBS (streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing) was used for sequence

read assessment and clean-up, alignment to reference genome, methylation status extraction, and CpG reporting and annotation. CpGs with low coverage (≤ 10) were excluded. Tertiary analysis consisted of removing non-informative or low sample coverage CpGs, and identifying methylated CpG regions with low background and dense clusters within sliding
5 100 bp windows. Read-depth criteria were based on the desired statistical power to detect a 10% difference in the %-methylation between cases and controls. Statistical significance was determined by logistic regression of the methylation percentage per DMR, based on read counts. To account for varying read depths across individual subjects, an over-dispersed
10 logistic regression model was used, where dispersion parameter was estimated using the Pearson Chi-square statistic of the residuals from fitted model. DMRs, ranked according to their significance level, were further considered if %-methylation in control groups was $\leq 1\%$ and $\geq 10\%$ in cancers. In most organ sites, this resulted in hundreds of potential candidates. Additional filters utilized were area under the receiver operating characteristic curve (AUC), %-methylation case/control fold change (FC), and positive sample to sample co-methylation
15 of CpGs throughout the DMR (and lack thereof in controls).

Technical and Biological Tissue Validation

Methylation specific PCR (MSP) marker assays were developed for 120 of the most promising DMRs from the PCa discovery dataset – as determined by the criteria listed above.
20 Primers were designed either by software (Methprimer - University of California, San Francisco CA) or by hand. Assays were rigorously tested and optimized by SYBR Green qPCR on bisulfite converted (methylated and unmethylated genomic DNA), unconverted, and non-template controls. Assays which cross-reacted with negative controls were either redesigned or discarded. In addition, melting curve analysis was performed to ensure specific
25 amplification was occurring. For the technical validation phase, the same samples used for the RRBS discovery were retested by qMSP. A β -actin assay designed to be methylation blind was used as a denominator representing total DNA copies. The data were analyzed by logistic regression and the AUC and signal to background results compared to the discovery values. Approximately 27% of the markers underperformed and were eliminated. The
30 remainder (N=72) were tested by qMSP on an expanded set of independent tissue samples. The results were analyzed logistically and outcome metrics were AUC, FC, and robust case sample % methylation.

Across Organ Validation

To assess how the best methylation markers performed outside the prostate, a comparative CpG %methylation matrix was constructed using the sequencing reads for the validation DMRs across the prostate samples compared to other major cancers previously sequenced – colon, pancreatic, esophageal, liver, and stomach. A final panel of markers was chosen to test in plasma based on 1) overall performance in the biological tissue validation phase and 2) the site specific characteristics of the markers across other cancers. To best detect PCa in blood, given the excess of non-PCa DNA, a robust marker panel was chosen that would exhibit both universal and prostate specific cancer signals.

QuARTs Assay Design and Plasma Validation

DNA was extracted from 3-4mL of banked frozen plasma by the following automated silica bead method (see, e.g., U.S. Patent Application Serial No. 15/335,111):

1	2ml of Te buffer (1mM Tris 0.1mM EDTA)
2	100µl of 120 cp/ul Zebra Fish in 0.4 ng/µl of Fish DNA diluent
3	7ml of plasma Lysis buffer (4.3M GTC 10% IGEPAL)
4	2ml of plasma
5	incubate at 55°C for 1 hour
6	Add 200µl binding beads
7	Add 2.8ml of 100% Isopropanol
8	Incubate at 30°C for 30 minutes.
9	magnetize the beads and remove the supernatant
10	Add 750µl 3M GuHCl 56.8% EtOH to resuspend binding beads
11	shake at 400 RPM for 2 minute.
12	Bind beads and aspirate supernatant to waste
13	1000µl wash 1 (80%ETOH), incubate at 30°C for 3 minutes, Bind beads and aspirate supernatant to waste
14	500µl wash 1 (80%ETOH), incubate at 30°C for 3 minutes, Bind beads and aspirate supernatant to waste
15	250µl wash 1 (80%ETOH), incubate at 30°C for 3 minutes, Bind beads and aspirate supernatant to waste

16	250µl wash 1 (80%ETOH), incubate at 30°C for 3 minutes, Bind beads and aspirate supernatant to waste
17	Dry at 70°C, 15 minutes, with shaking.
18	Add 125ul of Te buffer (1mM Tris 0.1mM EDTA) incubate 65°C for 25 minutes with shaking
19	Bind beads and transfer supernatent containing DNA to clean tubes
20	Store at -20°C until use.

The DNA was then bisulfite converted and purified, using the following method:

1	5ul 0.36% BSA
2	70ul of Sample
3	5ul 1.6N NaOH
4	Incubate (denaturation) 20' @ 42°C
5	cool for 8'
6	add 120ul of the Ammonium Bisulfite
7	Incubate (conversion) 75' @ 65°C (shake 3')
8	add 750ul of the 7M GuHcl
9	add 50ul binding beads
10	Incubate for 30' @ 30°C with shaking
11	Bind the beads
12	Aspirate supernatent to waste
13	add 1000ul 80% ETOH
14	Incubate for 3' @ 30°C with shaking
15	Bind the beads
16	Aspirate supernatent to waste
17	Dispense 200ul Desulphonation solution
18	Incubate for 7' @ 30°C with shaking
19	Bind the beads
20	Aspirate supernatent to waste
21	add 250ul 80% ETOH
22	Incubate for 3' @ 30°C with shaking

23	Bind the beads
24	Aspirate supernatant to waste
25	Dry beads for 15' @ 70 ⁰ C with shaking
26	Add 80ul of Te buffer (1mM Tris 0.1mM EDTA)
27	incubate 65°c for 25 minutes with shaking
28	Bind beads and transfer supernatant containing DNA to clean tubes
29	Store at -20°C until use.

Samples (10uL) were then run on an ABI real-time PCR instrument in the QuARTs-X (see, e.g., U.S. Patent Application Serial No. 15/335,096) format using primers and probes developed from the DMR sequence. Plasmids containing the marker sequence of interest were obtained from Genscript and diluted in 1X QuARTs reagents to a nominal concentration of 1 copy per 15 ul reaction. The reaction mix was distributed to each of 96 wells, cycled for 45 cycles on a LightCycler, and data collected. Wells were given a call of either containing or not containing sample. The Poisson random variable was set at 1 and average rate of success values were entered by trial and error and used to calculate the cumulative probability for that value. When the cumulative probability equals the percent of wells with signal, the correct average rate of success, in this case copy number, has been found. These plasmids were diluted and used as assay standards.

QuARTs-X was performed by first creating a pre-amplification plate of samples performed with primers for up to 12 targets that undergo 11 cycles of amplification. This product was then diluted 1:9 and used as template for subsequent QuARTs reactions that contain only three targets in a triplex reaction. Standards used to calculate strand counts did not go through the pre-amplification. By pre-amplifying the samples but not the standards the sensitivity of the assays was increased.

Results were analyzed by regressive partitioning (rPart). Using logistic regression to combine multiple methylation markers into a single risk score is a standard technique. However, it is difficult to discover and/or model high order interactions between markers within the logistic model. This limits the prediction capabilities of the panel of markers when such effects exist. Regression partitioning trees (rPart) is a decision tree approach that is able to discover high order interactions between the markers in such a way as to maximize the predictive accuracy of a panel of markers.

Example II.

This example describes the RRBS results and technical validation results.

PCa yielded large numbers of discriminate DMRs, many of which have not been identified before. Comparing the methylation of PCa samples to normal prostate, 256 regions were identified which met the AUC >0.85, FC >20, and p-value <0.05 cut-offs. 22 of these regions had an AUC of 1. When comparing PCa and normal prostate methylation to that of buffy coat samples, 1895 regions were above the cut-offs, with 827 having a perfect AUC score. FCs in both comparisons extended into the hundreds and thousands, respectively. Potential DMRs which differentiated Gleason 7+ PCa (aggressive, treatment indicated cancer) vs Gleason 6 PCa (indolent in most cases, treatment usually not required) were searched. 129 DMRs were observed with FC >2 (7+/6) with the highest FC = 72.

The second phase in the biomarker development process was to address the uncertainty arising from the relatively small sample size in the initial discovery phase. Re-testing the same samples using a different technology platform on a smaller number of DMR or marker candidates was a first step toward this end. Real-time or quantitative methylation specific PCR (qMSP) using SYBR Green is an easy to use method which has high analytical sensitivity and specificity.

DMRs were selected by taking top candidates from all three comparisons by increasing cut-offs until a manageable number of regions (N=120) were obtained (see, Table 1).

Table 1.

DMR No.	Gene Annotation	Chromosome No.	Region on Chromosome (starting base-ending base)
1	ADCY9	16	4165628-4165833
2	AKR1B1_3298	7	134143071-134143542
3	AKR1B1_3644	7	134143644-134143716
4	ANKRD35	1	145562809-14562898
5	CLIP4	2	29338053-29338117
6	CYBA_7733	16	88717482-88717805
7	FLJ45983	10	8097100-8097859
8	GRASP	12	52400932-52401144

9	GSTP1	11	67351212-67351638
10	HAPLN3	15	89438198-89438734
11	HCG4P6	6	29894504-29894683
12	HES5_0822	1	2460822-2460998
13	HES5_1047	1	2461047-2461876
14	ITPRIPL1	2	96990982-96991303
15	LRRC4	7	127671395-127672250
16	MAX.chr2.97193166-97193253	2	97193065-97193253
17	MAX.chr3.193	3	193776187-193776257
18	MAX.chr3.72788028-72788112	3	72788028-72788206
19	PTPRN2	7	157361654-157361753
20	RAI1_7469	17	17627094-17628093
21	RASSF2	20	4803273-4803687
22	Septin9_0074	17	75370074-75370187
23	Septin9_0492	17	75370492-75370581
24	SLC43A3	11	57194414-57194645
25	SLC4A11	20	3218937-3219001
26	SMAD3	15	67413704-67413772
27	SSBP4	19	18539756-18540408
28	TJP2	9	71788646-71789457
29	ABLIM1	10	116391692-116391769
30	AOX1	2	201450664-201450868
31	C3orf18	3	50604997-50605357
32	EIF5A2	3	170625931-170626391
33	EPSTI1	13	43566083-43566214
34	FBXO30	6	146136383-146136441
35	FLOT1_1586	6	30711586-30711681
36	FLOT1_1767	6	30711767-30711864
37	FLOT1_1904	6	30711904-30711966
38	FOSL1	11	65666866-65667056
39	GAS6	13	114566059-114566687
40	GSDMD	8	144640720-144640772
41	KCNK4	11	64059874-64059994

42	LOC100129726_1716	2	43451464-43452664
43	MAX.chr10.74079656-74079694	10	74079656-74079861
44	SERPINB9_3389	6	2903389-2903705
45	SLCO3A1	15	92396091-92396343
46	TPM4_7473	19	16187401-16187514
47	TPM4_8047	19	16187580-16188154
48	AGPS_7349	2	178257349-178257413
49	AGPS_7497	2	178257497-178257568
50	AGPS_7696	2	178257696-178257732
51	AMPD3	11	10472267-10472338
52	CH25H	10	90967594-90967655
53	GALR3	22	38214733-38214808
54	HDAC7_6722	12	48206722-48206761
55	MAX.chr1.227976339-227976430	1	227976339-227976430
56	MAX.chr1.8014263-8014319	1	8014263-8014319
57	MAX.chr19.34311051-34311120	19	34311051-34311120
58	NCRNA00092_4149	9	98784149-98784195
59	ZNF655_6084	7	99156084-99156145
60	ZNF655_6545	7	99156545-99156606
61	ZNF655_6762	7	99156762-99156852
62	ABCB1	7	87229774-87229851
63	ACOXL	2	111875299-111875585
64	ANXA2	15	60690904-60690949
65	APBB1IP	10	26727818-26728120
66	ARPC1B_1906	7	98971906-98971950
67	ARPC1B_1967	7	98971967-98971997
68	ATP2B4	1	203598589-203598782
69	CHST11_2032	12	104852032-104852137
70	CHST11_2206	12	104852206-104852307
71	DOK1_2325	2	74782325-74782452

72	FLJ34208	3	194208259-194208471
73	HCG4P6	6	29894504-29894683
74	HEBP2	6	138724600-138724667
75	HEYL	1	40105264-40105646
76	KLF16	19	1855656-1855656
77	LAMA3	18	21269864-21270082
78	LIME1	20	62369129-62369383
79	LOC339674	22	42353799-42353881
80	LOC440925	2	171570371-171570463
81	MAX.chr1.61519554-61519667	1	61519406-61519667
82	MAX.chr1.61519679-61519759	1	61519679-61519759
83	MAX.chr12.48592041-48592162	12	48592041-48592162
84	MAX.chr17.77786640-77786733	17	77786640-77786733
85	PDE4D	5	58334709-58335864
86	PLCL2	3	16925808-16925889
87	SIM2	21	38119920-38120410
88	STX16	20	57224816-57225220
89	WNT3A	1	228225487-228225688
90	ABHD15	17	27893168-27893592
91	ADAP1	7	963082-963154
92	ADD3	10	111767345-111767579
93	AXIN1	16	374825-375308
94	BCL2L11_6887	2	111876440-111876914
95	BCL2L11_6935	2	51717908-51718147
96	BIN2_7908	12	51717908-51717961
97	CCDC88C	14	91790497-91790556
98	CTBP1	4	1210660-1210752
99	DOK1_2096	2	74782096-74782223
100	DOK1_2234	2	74782234-74782302
101	DOK1_2475	2	74782475-74782572
102	FAM129C	19	17634139-17634203
103	FAM78A_1379	9	134151379-134151451

104	FAM78A_8684	9	134148528-134148765
105	FNBP1	9	132650777-132650834
106	GNG7_1972	19	2561972-2562075
107	GNG7_2119	19	2562119-2562198
108	HMHA1_9304	19	1069304-1069391
109	INPP5D	2	233925169-233925288
110	LYL1	19	13210124-13210498
111	MAX.chr15.95128172-95128228	15	95128172-95128228
112	MAX.chr16.11327022-11327151	16	11327022-11327292
113	MAX.chr16.50308415-50308535	16	50308415-50308535
114	MAX.chr4.1049805-1049912	4	1049805-1049912
115	MAX.chr9.134128109-134128241	9	134128109-134128241
116	NCOR2	12	124950727-124950905
117	OSM	22	30662704-30662800
118	S1PR4_0092	19	3180092-3180379
119	S1PR4_8637	19	3178410-3178763
120	S1PR4_9843	19	3179843-3180058

DMR sequences also had to exhibit significant co-methylation or contiguous methylation throughout the addressed CpGs on a per strand basis. qMSP and other amplification-based methods work best when all addressed CpGs are methylated (in cases) and unmethylated (in controls). After QC testing on standards (bisulfite treated universal methylated genomic DNA) and control samples (bisulfite treated unmethylated genomic DNA, non-converted genomic DNA, etc.), 99 regions performed with sufficient linearity, specificity, and robustness – and were used to re-test the phase 1 samples. Logistically analyzed results for most assays were comparable to the %methylation derived numbers from the sequencing phase. Total DNA strands in all samples were in excess of 100, with an average in the 1000s. The Z-markers (see, e.g., U.S. Patent Application Serial No. 14/966,617) continued to exhibit AUCs of 1 and extremely high FCs compared to normal buffy coat samples. When the cancer to benign ratios for the Z-marker candidates were investigated, about half were 1 to 1, with the rest having ratios between 2 and 10 (mid-Z). Of

the cancer vs benign markers, 34 had AUCs in the 0.95 to 1 range. All were negative with respect to buffy coat samples. Markers which differentiated Gleason 7+ cancers from Gleason 6 cancers in the phase 1 results continued their performance, in general, in the validation test. 26 markers had FCs in excess of 2 with the highest at 292.

- 5 Table 2 shows for the DMRs identified in Table 1 1) the area under the curve (AUC) for Prostate Cells with Gleason score at or over 6 vs. Benign Prostate Cells, 2) the Fold Change (FC) for Prostate Cells with Gleason score at or over 6 vs. Benign Prostate Cells, and 3) the Fold Change (FC) for Prostate Cells with Gleason score at or over 6 vs. Buffy (Norm).

10 Table 2.

DMR No.	Gene Annotation	AUC for Prostate Cells with Gleason at or over 6 vs. Benign Prostate Cells	FC for Prostate Cells with Gleason at or over 6 vs. Benign Prostate Cells	FC for Prostate Cells with Gleason at or over 6 vs. Buffy (Norm)
1	ADCY9	0.951	15.81	64.80
2	AKR1B1_3298	0.9857	24.80	108.96
3	AKR1B1_3644	0.9948	20.00	149.37
4	ANKRD35	0.9889	10.29	74.38
5	CLIP4	0.9905	14.31	68.96
6	CYBA_7733	0.9873	10.13	100.86
7	FLJ45983	0.9575	10.02	55.98
8	GRASP	1	65.34	54.18
9	GSTP1	1	29.28	54.11
10	HAPLN3	0.9984	27.21	83.77
11	HCG4P6	0.9531	18.31	69.37
12	HES5_0822	0.9698	13.91	74.47
13	HES5_1047	0.9714	18.20	63.37
14	ITPRIPL1	1	33.50	146.96
15	LRRC4	0.9802	244.30	23.20
16	MAX.chr2.97193166-97193253	0.9968	14.23	94.29
17	MAX.chr3.193	0.9762	18.09	80.44
18	MAX.chr3.72788028-72788112	0.981	19.72	64.91
19	PTPRN2	0.9841	28.83	57.62

20	RAI1_7469	0.9984	28.53	50.04
21	RASSF2	0.9921	12.47	163.35
22	Septin9_0074	0.9524	11.13	52.90
23	Septin9_0492	0.9627	29.96	58.55
24	SLC43A3	0.9546	10.74	62.40
25	SLC4A11	0.9921	12.12	57.37
26	SMAD3	1	17.07	87.91
27	SSBP4	0.9841	31.78	164.83
28	TJP2	0.9849	26.83	79.08
29	ABLIM1	0.9873	13.10	11.21
30	AOX1	0.9889	56.18	27.46
31	C3orf18	0.9429	11.06	33.72
32	EIF5A2	0.9635	12.42	26.45
33	EPSTI1	0.9429	13.68	40.17
34	FBXO30	0.946	31.76	87.79
35	FLOT1_1586	0.9968	8.95	57.06
36	FLOT1_1767	0.981	12.92	31.67
37	FLOT1_1904	0.9984	15.87	36.70
38	FOSL1	0.9762	13.20	27.90
39	GAS6	1	31.31	41.78
40	GSDMD	0.9635	15.23	53.83
41	KCNK4	0.9905	9.67	17.52
42	LOC100129726_1716	0.9714	31.51	114.94
43	MAX.chr10.74079656-74079694	0.9603	16.06	52.28
44	SERPINB9_3389	0.9841	13.91	54.70
45	SLCO3A1	0.9619	10.74	26.11
46	TPM4_7473	0.9832	36.60	269.51
47	TPM4_8047	0.9911	29.55	70.39
48	AGPS_7349		9.80	279.97
49	AGPS_7497		13.66	81.68
50	AGPS_7696		3.23	60.01
51	AMPD3	0.7757	29.14	115.79
52	CH25H	0.6771	6.67	311.57
53	GALR3	0.6818	5.85	50.02
54	HDAC7_6722	0.7411	10.06	50.66
55	MAX.chr1.227976339-227976430	0.7386	10.84	61.80
56	MAX.chr1.8014263-8014319	0.7556	17.68	449.60
57	MAX.chr19.34311051-34311120	0.6258	8.23	79.12

58	NCRNA00092_4149	0.6765	5.17	79.44
59	ZNF655_6084	0.8301	4.72	166.97
60	ZNF655_6545	0.8833	20.32	145.27
61	ZNF655_6762	0.7778	6.34	68.68
62	ABCB1	1	6.76	81.06
63	ACOXL	1	3.52	117.30
64	ANXA2	1	12.61	152.60
65	APBB1IP	1	1.41	112.40
66	ARPC1B_1906	1	2.12	295.90
67	ARPC1B_1967	1	1.45	228.10
68	ATP2B4	1	4.67	117.80
69	CHST11_2032	1	18.64	228.90
70	CHST11_2206	1	29.27	201.70
71	DOK1_2325	1	2.03	99.53
72	FLJ34208	1	2.81	137.00
73	HCG4P6	1	4.27	160.20
74	HEBP2	1	3.99	72.73
75	HEYL	1	5.64	71.91
76	KLF16	1	0.97	92.50
77	LAMA3	1	8.87	41.75
78	LIME1	1	2.98	112.30
79	LOC339674	1	2.48	52.99
80	LOC440925	1	1.69	187.50
81	MAX.chr1.61519554-61519667	1	2.71	131.00
82	MAX.chr1.61519679-61519759	1	2.32	157.20
83	MAX.chr12.48592041-48592162	1	8.24	73.55
84	MAX.chr17.77786640-77786733	1	8.18	42.27
85	PDE4D	1	3.85	91.96
86	PLCL2	1	3.25	101.40
87	SIM2	1	2.45	72.63
88	STX16	1	1.76	63.07
89	WNT3A	1	2.26	55.46
90	ABHD15	1	1.18	262.70
91	ADAP1	1	1.23	251.70
92	ADD3	1	38.62	342.40
93	AXIN1	1	1.08	499.60
94	BCL2L11_6887	1	1.39	325.50
95	BCL2L11_6935	1	1.42	274.50
96	BIN2_7908	1	1.07	413.00

97	CCDC88C	1	1.42	404.30
98	CTBP1	1	1.16	606.10
99	DOK1_2096	1	2.84	272.00
100	DOK1_2234	1	2.30	280.80
101	DOK1_2475	1	1.39	280.00
102	FAM129C	1	0.97	255.10
103	FAM78A_1379	1	1.03	899.10
104	FAM78A_8684	1	1.00	524.10
105	FNBP1	1	1.01	537.80
106	GNG7_1972	1	1.09	689.20
107	GNG7_2119	1	1.11	1503.00
108	HMHA1_9304	1	0.91	261.30
109	INPP5D	1	1.01	526.70
110	LYL1	1	1.07	790.10
111	MAX.chr15.95128172-95128228	1	1.13	330.20
112	MAX.chr16.11327022-11327151	1	1.48	656.90
113	MAX.chr16.50308415-50308535	1	0.98	407.50
114	MAX.chr4.1049805-1049912	1	1.26	362.60
115	MAX.chr9.134128109-134128241	1	1.01	356.90
116	NCOR2	1	0.94	438.30
117	OSM	1	1.07	504.30
118	S1PR4_0092	1	1.00	453.00
119	S1PR4_8637	1	1.02	461.40
120	S1PR4_9843	1	1.03	575.00

Example III.

Best performing candidate markers identified through the experiments described in Example II were selected for qMSP biological validation on DNA extracted from independent archival case and control tissues.

73 markers were selected (see, Table 3) from the phase 2 study to run on an independent set of prostate tissue (N = 35 normal prostate, 19 Gleason score 6, 31 Gleason score 7+) and normal buffy coat (N=36) samples. The 27 markers which were eliminated either had sub 0.85 AUCs – mainly in the cancer vs benign set, or less than perfect positive methylation in the Z-marker set. Most of the Gleason 7+ vs 6 markers were carried forward. All samples were assayed by qMSP, as before. DMR genomic coordinates and AUC for

Gleason score 7+ prostate tissue versus normal benign prostate tissue for the 73 assays are listed in Table 3, and respective primer sequences provided in Table 4.

Table 3.

Marker	Chr	Strand	AUC (7+/normal)	DMR No.
ABCB1	7	RS	0.91493	62
ABLIM1	10	FS	0.88976	29
ACOXL	2	FS	0.9592	63
ADCY9	16	RS	0.86372	1
ADD3	10	FS	0.92361	92
AGPS_7349	2	FS	0.69705	48
AGPS_7497	2	FS	0.6849	49
AGPS_7696	2	FS	0.75174	50
AKR1B1_3298	7	RS	0.92622	2
AKR1B1_3644	7	RS	0.96732	3
ANKRD35	1	FS	0.94618	4
ANXA2	15	RS	0.96962	64
ARPC1B_1906	7	FS	0.84028	66
ARPC1B_1967	7	FS	0.86024	67
AXIN1	16	RS	0.63411	93
BIN2_7908	12	RS	0.58854	96
CHST11_2206	12	RS	0.97861	70
CTBP1	4	FS	0.46007	98
EIF5A2	3	FS	0.92014	32
FAM78A_8684	9	RS	0.57813	104
FBXO30	6	FS	0.82899	34
FLJ45983	10	FS	0.99049	7
FLOT1_1586	6	FS	0.94355	35
FLOT1_1904	6	FS	0.92929	37
FNBP1	9	FS	0.61198	105
FOSL1	11	FS	0.89583	38
GALR3	22	RS	0.78559	53
GAS6	13	FS	0.98099	39
GNG7_2119	19	RS	0.75955	107
GRASP	12	RS	0.96732	8
GSDMD	8	FS	0.93576	40
GSTP1	11	FS	0.94792	9
HAPLN3	15	FS	0.95781	10
HCG4P6	6	FS	0.9836	11
HDAC7_6722	12	RS	0.81858	54
HEBP2	6	RS	0.93403	74
HES5_0822	1	RS	0.95247	12
HES5_1047	1	RS	0.94097	13
ITPRIPL1	2	RS	0.96078	14
KCNK4	11	FS	0.9798	41

KLF16	19	RS	0.53993	76
LAMA3	18	RS	0.87413	77
LOC100129726_1716	2	RS	0.87153	42
LOC339674	22	FS	0.90885	79
LOC440925	2	FS	0.82726	80
LRRC4	7	RS	0.94593	15
LYL1	19	RS	0.56597	110
MAX.chr1.61519554-61519667	1	RS	0.95486	81
MAX.chr1.8014263-8014319	1	FS	0.80729	56
MAX.chr10.74079656-74079694	10	RS	0.86372	43
MAX.chr12.48592041-48592162	12	RS	0.9321	83
MAX.chr15.95128172-95128228	15	FS	0.58681	111
MAX.chr16.11327022-11327151	16	FS	0.81684	112
MAX.chr17.77786640-77786733	17	RS	0.78125	84
MAX.chr2.97193166-97193253	2	FS	0.98099	16
MAX.chr3.193	3	FS	0.97683	17
MAX.chr3.72788028-72788112	3	FS	0.9798	18
NCRNA00092_4149	9	FS	0.84722	58
PDE4D	5	RS	0.8776	85
RAI1_7469	17	RS	0.97656	20
RASSF2	20		0.97861	21
S1PR4_9843	19	FS	0.52778	120
Septin9_0492	17	FS	0.93142	23
SERPINB9_3389	6	FS	0.96019	44
SIM2	21	RS	0.81343	87
SLC4A11	20	FS	0.95399	25
SLCO3A1	15	FS	0.62448	45
SSBP4	19	FS	0.9401	27
STX16	20	FS	0.87153	88
TPM4_8047	19	RS	0.95722	47

WNT3A	1	FS	0.84462	89
ZNF655_6084	7	FS	0.94271	59
ZNF655_6545	7	FS	0.80339	60

Table 4.

Marker	Primer Sequence	DMR No.
ABCB1	Left M primer TTGTTTTTTGAGTTCGCGGGC (SEQ ID NO: 99)	62
	Right M primer ACCAATACGATTCTCCCTCCCGAT (SEQ ID NO: 100)	
ABLIM1	Left M primer TTTCGACGAGTAGGATTGAAGAAGGAACG (SEQ ID NO: 127)	29
	Right M primer GCGAATCTATCTACCGAAACGCGCT (SEQ ID NO: 128)	
ACOXL	Left M primer AGTTAAGTTTTAACGGGTGTGGCGG (SEQ ID NO: 93)	63
	Right M primer AAACGTCGATAAAACGAACGTCGTA (SEQ ID NO: 94)	
ADCY9	Left M primer TTTCGGGCGTTTTAGGTTTCGTTTC (SEQ ID NO: 25)	1
	Right M primer GACTCAACGATACTCCCACCGCC (SEQ ID NO: 26)	
ADD3	Left M primer CGAGTTGTATAGTTAGAAGAGGACGT (SEQ ID NO: 1)	92
	Right M primer AACCGAAAAACCTAATTCGAAACG (SEQ ID NO: 2)	
AGPS_7349	Left M primer GGGGTAGAGAATGTGAAGTTTTAGACGT (SEQ ID NO: 63)	48
	Right M primer ACCGCGACGACTTAACGACG (SEQ ID NO: 64)	
AGPS_7497	Left M primer TTTTTTATTCGCGTTTAGCGGTTTCG (SEQ ID NO: 65)	49
	Right M primer CCGCCATAACTACCGCCTTATACTACCG (SEQ ID NO: 66)	
AGPS_7696	Left M primer TTAATGAGTGTAAGCGCGGAGAGTCG (SEQ ID NO: 67)	50

	NO: 67)	
	Right M primer CGAAAATAACCGTAAACGCTACCGT (SEQ ID NO: 68)	
AKR1B1_3298	Left M primer GATTTCTGGGCGTAGATAGGGACGT (SEQ ID NO: 143)	2
	Right M primer ACAACCGAACTACAAATACCTCGAA (SEQ ID NO: 144)	
AKR1B1_3644	Left M primer GGGGTTCGTTTTATATTTTCGCGC (SEQ ID NO: 27)	3
	Right M primer CAAAATCACTCGAAATCCCTCGCC (SEQ ID NO: 28)	
ANKRD35	Left M primer GGGAGGTAGTTAGTTTACGGTAATACGA (SEQ ID NO: 29)	4
	Right M primer CTAAACCACCAACGAACCCCGAA (SEQ ID NO: 30)	
ANXA2	Left M primer GGGTTTAAAGTTATGGGTTTTATTTCGT (SEQ ID NO: 89)	64
	Right M primer TAAAACATCTCTACGCGACCTCGTA (SEQ ID NO: 90)	
ARPC1B_1906	Left M primer AGGGAGTTTTTCGTTGGTTGTCGAC (SEQ ID NO: 81)	66
	Right M primer ATACTACGAACTCCGCGCTCACACG (SEQ ID NO: 82)	
ARPC1B_1967	Left M primer AGGGAGTTTTTCGTTGGTTGTCGAC (SEQ ID NO: 83)	67
	Right M primer ATACTACGAACTCCGCGCTCACACG (SEQ ID NO: 84)	
AXIN1	Left M primer GGGGTATTAGTTTTTTATGAGATTGCGT (SEQ ID NO: 3)	93
	Right M primer TAAAAAATCCACCTATCGCTCGAA (SEQ ID NO: 4)	
BIN2_7908	Left M primer TTGTTAATTTTTTGGGGGTCGTCGT (SEQ ID NO: 5)	96
	Right M primer GAAAACCCGCACTTCCTCCTCGA (SEQ ID NO: 6)	

CHST11_2206	Left M primer TTTTITTAGTTTAGATTTCGGACGT (SEQ ID NO: 85)	70
	Right M primer TAAAAATAAACCCCATTCCTCCGAT (SEQ ID NO: 86)	
CTBP1	Left M primer TATAGGATTTTAAGTTCGCGAACGT (SEQ ID NO: 7)	98
	Right M primer TCAACCTACTTCCTTCCTATATAACGAT (SEQ ID NO: 8)	
EIF5A2	Left M primer ATCGTTTTATCGTAGAAGTCGGCGA (SEQ ID NO: 135)	32
	Right M primer TACGACCTAAACTAAATCCCCCGCA (SEQ ID NO: 136)	
FAM78A_8684	Left M primer GGAGTTAGAAGTTTTTTGGGAGGGGC (SEQ ID NO: 9)	104
	Right M primer AAACACGTAAACCCTTCTACCCGAC (SEQ ID NO: 10)	
FBXO30	Left M primer GTTTTTTCGTCGGTTAATTTAGCGT (SEQ ID NO: 141)	34
	Right M primer AAAAATAACGAATCACAACACCGTC (SEQ ID NO: 142)	
FLJ45983	Left M primer TAGTCGAGGTTATGGAGGTGACGGC (SEQ ID NO: 31)	7
	Right M primer ACTACCCGTAAACACGACGAA (SEQ ID NO: 32)	
FLOT1_1586	Left M primer TGTTTCGGAAGTTTTAGTTGGGGATACGT (SEQ ID NO: 121)	35
	Right M primer AACACCAAACGTACCTAATACGCGAA (SEQ ID NO: 122)	
FLOT1_1904	Left M primer GTTATTAGGATTTGGTAGAAGACGAI (SEQ ID NO: 119)	37
	Right M primer AACTACCAATCGAAAAACCGAA (SEQ ID NO: 120)	
FNBP1	Left M primer GCGTGATTGATGGGTGTATTACGT (SEQ ID NO: 11)	105
	Right M primer ATAACTTCCGATCCCTACAACGAA (SEQ ID NO: 12)	

	12)	
FOSL1	Left M primer GTTAGGAAGGGAGGGACGTTTCGG (SEQ ID NO: 131)	38
	Right M primer CGAAAACTACGAACACGTATCGAC (SEQ ID NO: 132)	
GALR3	Left M primer TG TAGTAGGATACGTTGAGTCGCGG (SEQ ID NO: 69)	53
	Right M primer GCGATAAACTCCACGCCGTT (SEQ ID NO: 70)	
GAS6	Left M primer TAGTTTAGTAGAGGGAGGGTCGCGG (SEQ ID NO: 117)	39
	Right M primer CGCGAAAAAACC GAAAATCCGTT (SEQ ID NO: 118)	
GNG7_2119	Left M primer TAGTTTCGGGGTTGCGGTGATTAC (SEQ ID NO: 13)	107
	Right M primer AAAATCCGAAACAAACATTACGCC (SEQ ID NO: 14)	
GRASP	Left M primer TGTTTTCGGATACGGCGAGC (SEQ ID NO: 33)	8
	Right M primer ACGAACGAACTATACGCGACGCT (SEQ ID NO: 34)	
GSDMD	Left M primer GTTCGTTTAGAAGGTTTCGCGTCGTATAC (SEQ ID NO: 145)	40
	Right M primer ACCTTTCCCGAAACCTAAACTTCCG (SEQ ID NO: 146)	
GSTP1	Left M primer TCGTGATTAGTATTGGGGCGGAGC (SEQ ID NO: 35)	9
	Right M primer GAAACTCCAACGAAAACCTCGCGAC (SEQ ID NO: 36)	
HAPLN3	Left M primer AAGCGGTAAGGGAGGAATTCGGTTC (SEQ ID NO: 37)	10
	Right M primer GACCCCCGAAACTCTAACCGTCG (SEQ ID NO: 38)	
HCG4P6	Left M primer GGATCGGAGTATTGGGATCGGAGTATAC (SEQ ID NO: 39)	11

	Right M primer AACTCTAATAATAATAACGACGCGAC (SEQ ID NO: 40)	
HDAC7_6722	Left M primer GTAGTATTTTATTTTATCGGGCGA (SEQ ID NO: 71)	54
	Right M primer AAAAAATCACACCTCCTTCAACGCT (SEQ ID NO: 72)	
HEBP2	Left M primer TTCGAGGTTTTTCGGGGCGAC (SEQ ID NO: 103)	74
	Right M primer CCCTCCTATCCGTAACTTTCGCGTA (SEQ ID NO: 104)	
HES5_0822	Left M primer GTGATTCGGCGGGATTGCG (SEQ ID NO: 41)	12
	Right M primer GACGAAAAACGCTTCCCTACAAACGA (SEQ ID NO: 42)	
HES5_1047	Left M primer AAGAGTTTGTATTAGGATTATAGCGA (SEQ ID NO: 43)	13
	Right M primer CGTAAACGTCAAAACTACACGAC (SEQ ID NO: 44)	
ITPRIPL1	Left M primer GGGATTTAGGGTTAGGTTATTTATCGT (SEQ ID NO: 45)	14
	Right M primer ACCGCGCTATCTCTTTAAATCGTA (SEQ ID NO: 46)	
KCNK4	Left M primer TTAGTTAGGAAGTAGGGCGAGGCGA (SEQ ID NO: 125)	41
	Right M primer AAAACCGAACAACGCAAAACGAA (SEQ ID NO: 126)	
KLF16	Left M primer TGAGATTTTCGGTTAAAGGAAGGGGTC (SEQ ID NO: 95)	76
	Right M primer ACCTACCTATACGCCTCCAAACGAT (SEQ ID NO: 96)	
LAMA3	Left M primer GTTTTTATTTGGATGGTGTGGTCGT (SEQ ID NO: 115)	77
	Right M primer CCCGAACTCTACTACAAATTAATCGAA (SEQ ID NO: 116)	
LOC100129726_1716	Left M primer TACGTCGTTTCGTTATTTAGATTTATAATTTTGTC	42

	(SEQ ID NO: 133)	
	Right M primer CCAAACCCTACTACTACTACTACTAACTACG (SEQ ID NO: 134)	
LOC339674	Left M primer TGGTGGATCGCGATTTTCGTAAGAC (SEQ ID NO: 111)	79
	Right M primer CGCCGAAAACCAAATTATCGCG (SEQ ID NO: 112)	
LOC440925	Left M primer CGGTGAGTATTTGCGGGTTTTTCGT (SEQ ID NO: 87)	80
	Right M primer AAATCGCCTCTCCCGAACGC (SEQ ID NO: 88)	
LRRC4	Left M primer TAATTTGCGGAGGTAGGCGACGG (SEQ ID NO: 47)	15
	Right M primer CAATACTCTTATATATTAACGCCGCT (SEQ ID NO: 48)	
LYL1	Left M primer TGTTTCGTTTAGTTATGAAGTATATCGG (SEQ ID NO: 15)	110
	Right M primer ACTTAATCGCGCAACAACCGCA (SEQ ID NO: 16)	
MAX.chr1.6151 9554-61519667	Left M primer GTCGCGTTTTTTTATTTGTCGTTTCGT (SEQ ID NO: 91)	81
	Right M primer ATAAAAATCGATTCTACCGCGTCGCC (SEQ ID NO: 92)	
MAX.chr1.8014 263-8014319	Left M primer AATACGCGACGGTTTCGTTTATTGC (SEQ ID NO: 73)	56
	Right M primer CATAACGTAAATCCACTTCCGACGAC (SEQ ID NO: 74)	
MAX.chr10.740 79656- 74079694	Left M primer GGTTGTAAGGGGGTTTGGGTACGC (SEQ ID NO: 139)	43
	Right M primer ATTTCGAAAAAAACGCCCGATACGA (SEQ ID NO: 140)	
MAX.chr12.485 92041- 48592162	Left M primer TCGCGTTGACGGTTTGTGACG (SEQ ID NO: 101)	83
	Right M primer AACCCCGATCCGAAAAACCGAA (SEQ ID NO: 102)	
MAX.chr15.951	Left M primer	111

28172-95128228	CGGTTATATTATAAGAAAAGGAAGTTTTCGT (SEQ ID NO: 17)	
	Right M primer GAAAACCCAACTACACACCGCT (SEQ ID NO: 18)	
MAX.chr16.11327022-11327151	Left M primer TTGGTTTTTATTAAGTTATGTGACGA (SEQ ID NO: 19)	112
	Right M primer TAAAATTCCAAAAACGATAACGCT (SEQ ID NO: 20)	
MAX.chr17.77786640-77786733	Left M primer GGGTGGATTTTCGGGCGTTATAAATC (SEQ ID NO: 113)	84
	Right M primer CAAAACGACTCCCCGCCGAA (SEQ ID NO: 114)	
MAX.chr2.97193166-97193253	Left M primer GGTTTTAGGGAAATATCGGCGT (SEQ ID NO: 49)	16
	Right M primer AACTCAAACCGAAAAAATAATTGAT (SEQ ID NO: 50)	
MAX.chr3.193	Left M primer AAAGGTTTAGTTAAAGATGGAATCGT (SEQ ID NO: 51)	17
	Right M primer CTCGCGACGAAAAAAACCCGAA (SEQ ID NO: 52)	
MAX.chr3.72788028-72788112	Left M primer AGGATTCGACGGAGTTATATTCGT (SEQ ID NO: 53)	18
	Right M primer TAACATAACCACCCAACTCTCCCCGAA (SEQ ID NO: 54)	
NCRNA00092_4149	Left M primer CGTTTAGGGGGTTTCGAGCGTAGC (SEQ ID NO: 75)	58
	Right M primer CCCTAATTCCATCCTAAACGAATCGAC (SEQ ID NO: 76)	
PDE4D	Left M primer AATTCGTAGGAAGTAGTCGGTCGT (SEQ ID NO: 97)	85
	Right M primer TACCCTCCAAATTACCCAAACCGCT (SEQ ID NO: 98)	
RAI1_7469	Left M primer TTATAGTAGTTCGTCGAAAATATAAAGTTTCGTC (SEQ ID NO: 55)	20
	Right M primer CCGAAAAACCCAAAAAAACCCG (SEQ ID NO: 56)	
RASSF2	Left M primer GTCGTTTACGGTATTTGTTTCGTTC (SEQ ID NO: 57)	21
	Right M primer	

	ATCGCTAAACCTCAACCTAACGTC (SEQ ID NO: 58)	
S1PR4_9843	Left M primer TGAGGTAAAGGGATAGTTTTCGCGG (SEQ ID NO: 21)	120
	Right M primer AACTACAACCTCAAATACTCCGCACGCT (SEQ ID NO: 22)	
Septin9_0492	Left M primer GGTTTTGC GTTTTGC GTTCGC (SEQ ID NO: 23)	23
	Right M primer CCATTATATAAACTTCCCCTTCGCC (SEQ ID NO: 24)	
SERPINB9_3389	Left M primer GGTTTTATTATTCGGTGGTAGTCGG (SEQ ID NO: 129)	44
	Right M primer ACACGAAAACGACGACAACGCT (SEQ ID NO: 130)	
SIM2	Left M primer GGTCGTAGTTCGGGAAGTTCGG (SEQ ID NO: 105)	87
	Right M primer ATTCGACGAAAACACCGCGC (SEQ ID NO: 106)	
SLC4A11	Left M primer TTCGTTGTTTGTGTTTCGTTGTGCG (SEQ ID NO: 59)	25
	Right M primer CCGAATCTAAAACGCTTTCCTCTCGTA (SEQ ID NO: 60)	
SLCO3A1	Left M primer GCGATTTTGTAGCGTTTGGATATCGA (SEQ ID NO: 137)	45
	Right M primer GTCGAAAACCCACGAACCGTT (SEQ ID NO: 138)	
SSBP4	Left M primer GTAGCGTCGGCGTAGAGCGTAGATC (SEQ ID NO: 61)	27
	Right M primer ATAACGAATCCCCGCGAAAATTCTG (SEQ ID NO: 62)	
STX16	Left M primer CGGGAAATTTTCGGAAAATATATACGT (SEQ ID NO: 107)	88
	Right M primer TACGAAATTCCAACAAAAAACCGAA (SEQ ID NO: 108)	
TPM4_8047	Left M primer TTTTTTTATTTTATTTTTTTCGTCGT (SEQ ID NO: 123)	47
	Right M primer	

	GACTTCTACTTACTTCCCTAACCGTT (SEQ ID NO: 124)	
WNT3A	Left M primer CGGGCGGTTTATACGTTTTTCGC (SEQ ID NO: 109)	89
	Right M primer AAAAAAATCCCCATTCAAACGCT (SEQ ID NO: 110)	
ZNF655_6084	Left M primer TTGCGAAACGAGTTTTCGAATTATGGAC (SEQ ID NO: 77)	59
	Right M primer CCCCGAATATAAATAACGACCCCGAA (SEQ ID NO: 78)	
ZNF655_6545	Left M primer TTGGTTTATTTATTTTCGCGGATCGA (SEQ ID NO: 79)	60
	Right M primer AAAACACGATCGCCGACTCCTAACG (SEQ ID NO: 80)	

AUCs overall were excellent, although somewhat less than in the earlier validation – which was to be expected in this expanded independent set. As shown in Table 3, 20 markers had AUC values (Gleason 7+ vs normal prostate) in the 0.95 – 0.99 range and FCs 17-164.

- 5 %methylation in buffy coat samples was negligible except for GRASP for which there was a single outlier.

Since clinical follow up data was available on the prostate cases, it was decided to explore the prognostic aspects of the epigenetic marker candidates. Using regressive partitioning (rPart), a mathematical method for discovering and/or modelling high order interactions between markers within the logistic model, five prognostic markers (FAM78A, WNT3A, GAS6, LOC100129726, and MAX.chr3.727) were selected. The risk grouping defined by methylated DNA markers added significant prognostic content in predicting progression-free survival relative to Gleason scoring ($p < 0.0001$) whereas Gleason scoring had no added value relative to methylated DNA marker risk grouping ($p = 0.2174$).

15

Example IV.

Additional experiments were conducted to identify markers capable of distinguishing between PCa Gleason score over 7 versus Gleason score at 6 within prostate tissue. Such experiments utilized the QuARTs-X (quantitative allele specific real time target and signal assays) (see, e.g., U.S. Patent Application Serial No. 15/335,096). Table 5 shows marker

20

sensitivity at 100% and fold-change for PCa Gleason 7+ versus Gleason 6 within prostate tissue (the oligo sequences are provided in Table 6).

Table 5

DMR No.	Marker	Sensitivity for Gleason over 7 vs. Gleason 6 in Prostate Tissue Sample	Fold-Change for Gleason over 7 vs. Gleason 6 in Prostate Tissue Sample
17	MAX.chr3.193	90%	1.26
12	HES5_0822	84%	1.5
45	SLCO3A1	70%	2
47	TPM4_8047	68%	1.42
83	MAX.chr12.48592041-48592162	86%	1.26
39	GAS6	86%	1.02
18	MAX.chr3.72788028-72788112	78%	1.37
3	AKR1B1_3644	66%	1.18
87	SIM2	34%	1.7
11	HCG4P6	86%	0.99
70	CHST11_2206	88%	1.7
8	GRASP	82%	1.59
44	SERPINB9_3389	74%	1.49
85	PDE4D	62%	1.08
14	ITPRIPL1	76%	2.16
16	MAX.chr2.97193166-97193253	82%	1.34
37	FLOT1_1904	86%	1.44

5

Table 6

MARKER	OLIGO TYPE	OLIGO NAME	SEQUENCE
SERPINB9	Primer	SERPINB9_FP	TTTATTTTAGTCGTGCGCGG (SEQ ID NO: 147)
	Primer	SERPINB9_RP	ACGCGACACGAAAACGAC (SEQ ID NO: 148)
	Probe	SERPINB9_Pb_A5	CCACGGACG CGACAACGCTAA/3C6/ (SEQ ID NO: 149)
FLOT1	Primer	FLOT1_FP	GTAGTGTTTTGAGTTTAAGTTGTTT CG (SEQ ID NO: 150)
	Primer	FLOT1_RP	AACACCAAACGTACCTAATACG (SEQ ID NO: 151)

	Probe	FLOT1_Pb_A1	CGCCGAGG GCGAAAACGTAT/3C6/ (SEQ ID NO: 152)
HCG4P6	Primer	HCG4P6_FP	CGGAGTATACGGAATATTAGGTTG (SEQ ID NO: 153)
	Primer	HCG4P6_RP	AATAATAACGACGCGACATAAACA (SEQ ID NO: 154)
	Probe	HCG4P6_Pb_A5	CCACGGACG GCGTATAGATTG/3C6/ (SEQ ID NO: 155)
CHST11_2206	Primer	CHST11_2206_FP	GCGTTCGAGGGCGTTTT (SEQ ID NO: 156)
	Primer	CHST11_2206_RP	AATCTAACTAAAAAAAAAACGAA ACTCGT (SEQ ID NO: 157)
	Probe	CHST11_2206_Pb_A1	CGCCGAGG TCGCTTCCTAAA/3C6/ (SEQ ID NO: 158)
MAX.chr12.485	Primer	MAX.chr12.485_FP	GAGGAAATAGGCGTGGTTG (SEQ ID NO: 159)
	Primer	MAX.chr12.485_RP	AAAACCGAAAAAACTAATTCGTC A (SEQ ID NO: 160)
	Probe	MAX.chr12.485_Pb_A5	CCACGGACG GCGTTGACGGTT/3C6/ (SEQ ID NO: 161)
GRASP	Primer	GRASP_FP	CGCGCGGTATAGTTGCG (SEQ ID NO: 162)
	Primer	GRASP_RP	ACCCCAACGAACGAAC (SEQ ID NO: 163)
	Probe	GRASP_Pb_A1	CGCCGAGG CTATACGCGACG/3C6/ (SEQ ID NO: 164)
GAS6	Primer	GAS6_FP	GATTTTGAGGTTAGTTTCGTG (SEQ ID NO: 165)
	Primer	GAS6_RP	GAAAAACAACGCTACTACCGC (SEQ ID NO: 166)
	Probe	GAS6_Pb_A5	CCACGGACG GCGCGGAGTTGG/3C6/ (SEQ ID NO: 167)
HAPLN3	Primer	HAPLN3_FP	GTTTCGTATATAGTTTTAGAAGTTT GCGA (SEQ ID NO: 168)
	Primer	HAPLN3_RP	CGTCGAAAACTACCTAAAAACGA T (SEQ ID NO: 169)
	Probe	HAPLN3_Pb_A1	CGCCGAGG TACGTAACCTAA/3C6/ (SEQ ID NO: 170)
SLCO3A1	Primer	SLCO3A1_FP	CGATTTTGTAGCGTTTGGATATCG (SEQ ID NO: 171)
	Primer	SLCO3A1_RP	TCGAAAACCCACGAACC (SEQ ID NO: 172)

	Probe	SLCO3A1_Pb_A5	CCACGGACG CGTTAATACCCC/3C6/ (SEQ ID NO: 173)
MAX.chr3.193	Primer	MAX.chr3.193_FP	TTAAAGGTTTAGTTAAAGATGGAA TCGT (SEQ ID NO: 174)
	Primer	MAX.chr3.193_RP	CTCGCGACGAAAAAACCC (SEQ ID NO: 175)
	Probe	MAX.chr3.193_Pb_A1	CGCCGAGG CGAACTCCCAAC/3C6/ (SEQ ID NO: 176)
MAX.chr3.727	Primer	MAX.chr3.727_FP	GTGGTTTTATTTTCGTTTCGTTTCG (SEQ ID NO: 177)
	Primer	MAX.chr3.727_RP	AACTAACGAATATAACTCCGTCG A (SEQ ID NO: 178)
	Probe	MAX.chr3.727_Pb_A1	CGCCGAGG GCGACGTTTCGAG/3C6/ (SEQ ID NO: 179)
HES5	Primer	HES5_FP	GCGAGAGGAGTAGGTTTCG (SEQ ID NO: 180)
	Primer	HES5_RP	AACCTACGAACGCGCGA (SEQ ID NO: 181)
	Probe	HES5_Pb_A5	CCACGGACG ACCCGACGACCA/3C6/ (SEQ ID NO: 182)
PDE4D	Primer	PDE4D_FP	GCGTACGGTCGCGTATT (SEQ ID NO: 183)
	Primer	PDE4D_RP	CAATAACTCGACGAAACGCG (SEQ ID NO: 184)
	Probe	PDE4D_Pb_A1	CGCCGAGG GCGAAATTCTAA/3C6/ (SEQ ID NO: 185)
SIM2	Primer	SIM2_FP	GTTTAGCGCGGGTTTTTCG (SEQ ID NO: 186)
	Primer	SIM2_RP	CCCGAACTTCCCGAACT (SEQ ID NO: 187)
	Probe	SIM2_Pb_A5	CCACGGACG GCGGTAGTGGTC/3C6/ (SEQ ID NO: 188)
TPM4	Primer	TPM4_FP	GGAGAAAGGCGGGCG GA (SEQ ID NO: 189)
	Primer	TPM4_RP	CGTTCCCAAAAACGCGA (SEQ ID NO: 190)
	Probe	TPM4_Pb_A1	CGCCGAGG ACGACGACGTAT/3C6/ (SEQ ID NO: 191)
LRRC4_HC C	Primer	LRRC4_HCC_FP	CGTTCGTTTCGTTTCGTTTTGG (SEQ ID NO: 192)
	Primer	LRRC4_HCC_RP	CCCCGCCCCCTCTAAAC (SEQ ID NO: 193)

	Probe	LRRC4_HCC_Pb_A5	CCACGGACG CGAACGAAACGA/3C6/ (SEQ ID NO: 194)
ITPRIPL1	Primer	ITPRIPL1_FP	GTAGTGGCGGTTAGGTCG (SEQ ID NO: 195)
	Primer	ITPRIPL1_RP	TCACCGTCAATATTAATAAAAACA CGA (SEQ ID NO: 196)
	Probe	ITPRIPL1_Pb_A1	CGCCGAGG GCGTTGTTTTTT/3C6/ (SEQ ID NO: 197)
MAX.chr chr2.971931 66- 97193253	Primer	MAX.chr2. chr2.97193166- 97193253_FP	GCGAGGTTGCGGTTTTG (SEQ ID NO: 198)
	Primer	MAX.chr2. chr2.97193166- 97193253_RP	CGAAAAAATAATTTCGATTAAACCAA AACGC (SEQ ID NO: 199)
	Probe	MAX.chr2. chr2.97193166- 97193253_Pb_A5	CCACGGACG CCGATATTTCCT/3C6/ (SEQ ID NO: 200)
AKR1B1_36 44	Primer	AKR1B1_FP	GTTTCGTTTTATATTTTCGCGCG (SEQ ID NO: 201)
	Primer	AKR1B1_RP	CCGAACGTCCGCGAAAC (SEQ ID NO: 202)
	Probe	AKR1B1_Pb_A1	CGCCGAGG CGAACTACTCAA/3C6/ (SEQ ID NO: 203)
FLJ45983	Primer	FLJ45983_FP	GGGCGCGAGTATAGTCG (SEQ ID NO: 204)
	Primer	FLJ45983_RP	CAACGCGACTAATCCGC (SEQ ID NO: 205)
	Probe	FLJ45983_Pb_A1	CGCCGAGG CCGTCACCTCCA/3C6/ (SEQ ID NO: 206)

Example V.

Experiments were conducted wherein candidate marker sequences were compared in-silico across a pan-cancer RRBS sequencing data set to gauge the degree of site specific methylation for each marker.

DNA methylation signatures have been shown to accurately predict the site of a tumor within the human anatomy. To better define organ site-related specificity, an in-silico CpG x sample matrix of RRBS derived methylation values for each of the 73 DMR/markers across the multiple cancers and organ tissues was constructed. These included prostate, liver, colorectal, pancreatic, lung, esophageal, stomach, and bile duct tissues. Locational specificity could be modelled by quantitative methylation differences between organ sites as well as the

degree of contiguous methylation (defined by pattern recognition) throughout the DMR. As shown in Table 7, 8 markers demonstrated prostate cancer only specificity – and on the other extreme, 11 markers were universal to all cancers and tissues. In between were clusters of specificity at the same or differing degrees such as prostate/liver, prostate/colon/liver, etc. A subset of prostate markers remain undefined due to missing DMR sequences in the aligned reads of the other cancers.

Table 7.

DMR No.	Marker	AUC for Prostate Tissue with Gleason over 7 vs. Benign Prostate Tissue	FC for Prostate Tissue with Gleason over 7 vs. Benign Prostate Tissue	FC for Prostate Tissue with Gleason at or over 6 vs. Buffy	% meth (7+)	Tissue Specificity
62	ABCB1	0.91493	16	3496	15	prostate/lung
29	ABLIM1	0.88976	39	51	5	no data
63	ACOXL	0.9592	18	59094554	37	prostate/lung /pancreas
1	ADCY9	0.86372	19	38	4	no data
92	ADD3	0.92361	2290	3075	20	prostate specific
48	AGPS_7349	0.69705	12	554	3	no data
49	AGPS_7497	0.6849	7	7534	2	no data
50	AGPS_7696	0.75174	58	52722	13	no data
2	AKR1B1_3298	0.92622	32	604	37	prostate/colon/HCC
3	AKR1B1_3644	0.96732	128	6850	37	prostate/colon/HCC
4	ANKRD35	0.94618	5	4184	48	universal
64	ANXA2	0.96962	34	36464	66	prostate/HCC
66	ARPC1B_1906	0.84028	5	45311	16	prostate specific
67	ARPC1B_1967	0.86024	8	5800	23	prostate specific
93	AXIN1	0.63411	1	16677	143	universal
96	BIN2_7908	0.58854	1	10778	90	universal
70	CHST11_2206	0.97861	145	41729	27	prostate specific
98	CTBP1	0.46007	1	1188	114	universal
32	EIF5A2	0.92014	32	7138	16	no data
104	FAM78A_868	0.57813	1	8026	55	universal

	4					
34	FBXO30	0.82899	13	1726	19	no data
7	FLJ45983	0.99049	35	1480	54	prostate/mild universal
35	FLOT1_1586	0.94355	11	15687	64	prostate specific
37	FLOT1_1904	0.92929	65	322	38	prostate specific
105	FNBP1	0.61198	1	78760	97	universal
38	FOSL1	0.89583	69	481	5	no data
53	GALR3	0.78559	31	34071	15	no data
39	GAS6	0.98099	75	62587	47	prostate specific
107	GNG7_2119	0.75955	1	19012	100	universal
8	GRASP	0.96732	78	37	30	prostate/colon/HCC/lung
40	GSDMD	0.93576	28	14825	20	no data
9	GSTP1	0.94792	21	361117	77	prostate/HCC
10	HAPLN3	0.95781	60	969498	38	prostate specific
11	HCG4P6	0.9836	30	73301	58	prostate/mild universal
54	HDAC7_6722	0.81858	149	11130	12	no data
74	HEBP2	0.93403	14	10185	21	prostate specific
12	HES5_0822	0.95247	26	8884	29	prostate specific
13	HES5_1047	0.94097	13	12636	28	prostate specific
14	ITPRIPL1	0.96078	45	12163	19	prostate/colon/HCC/lung
41	KCNK4	0.9798	34	682	29	no data
76	KLF16	0.53993	1	1155	21	prostate/HCC
77	LAMA3	0.87413	79	106	18	prostate specific
42	LOC100129726_1716	0.87153	11	2594	16	no data
79	LOC339674	0.90885	14	447	29	prostate/HCC/lung (small cell only)
80	LOC440925	0.82726	3	2215	48	prostate/lung
15	LRRC4	0.94593	61	19639	48	prostate/colon/lung/pancreas
110	LYL1	0.56597	1	23253	59	no data
81	MAX.chr1.61519554-61519667	0.95486	5	5254	58	prostate/lung
56	MAX.chr1.8014263-8014319	0.80729	25	1770	10	prostate/lung

43	MAX.chr10.7 4079656- 74079694	0.86372	50	278757	14	no data
83	MAX.chr12.4 8592041- 48592162	0.9321	26	364	63	prostate specific
111	MAX.chr15.9 5128172- 95128228	0.58681	1	1363	104	universal
112	MAX.chr16.1 1327022- 11327151	0.81684	3	31278	324	universal
84	MAX.chr17.7 7786640- 77786733	0.78125	16	67	7	prostate/lung (small cell only)
16	MAX.chr2.97 193166- 97193253	0.98099	17	967	47	prostate/colo n/HCC/lung
17	MAX.chr3.19 3	0.97683	43	1605	66	prostate specific
18	MAX.chr3.72 788028- 72788112	0.9798	87	633	52	prostate/HC C
58	NCRNA0009 2_4149	0.84722	11	1101	21	no data
85	PDE4D	0.8776	4	835	59	prostate specific
20	RAI1_7469	0.97656	28	320	56	universal
21	RASSF2	0.97861	24	1528628	44	prostate/colo n/HCC
120	S1PR4_9843	0.52778	1	722	86	universal
23	Septin9_049 2	0.93142	21	3348	33	no data
44	SERPINB9_3 389	0.96019	164	3993	50	prostate specific
87	SIM2	0.81343	3	323	51	prostate specific
25	SLC4A11	0.95399	20	768	16	prostate/colo n/HCC
45	SLCO3A1	0.62448	4	6712	71	no data
27	SSBP4	0.9401	34	1043	25	prostate/colo n/HCC
88	STX16	0.87153	4	279	34	prostate/colo n
47	TPM4_8047	0.95722	23	2578	31	prostate specific
89	WNT3A	0.84462	5	2144	28	prostate/HC C
59	ZNF655_608 4	0.94271	34	(zero in denominat or)	11	no data
60	ZNF655_654 5	0.80339	5	57988	14	no data

Example VI.

Experiments were conducted wherein a set of high performing PCa markers were chosen for testing in blinded independent plasma samples to assess PCa detection in a clinical medium.

Twenty-five DMRs/markers (see, Table 8) which exhibited the best combination of performance metrics suited to analyte detection in complex biological media (e.g., blood plasma) were chosen for development of multiplex QuARTs-X (quantitative allele specific real time target and signal assays) – a testing platform frequently used with other cancer plasma studies. After initial in-silico tests on the uniqueness of the DMR sequences, design filters, and QC testing on pooled plasma controls, 17 designs were carried forward for testing on retrospectively collected frozen plasma samples from the Mayo prostate cancer biobank. 9 tri-plexes and 1 duplex reaction were developed, each of which included a control β -actin assay not affected by methylation. Two additional process controls were also tested. The QuARTs-X assays are listed in Table 8 (see Table 6 for the primer and probe information). Final marker strands (copies) were normalized to the β -actin control and expressed as % methylation. The top 4 markers were max.chr3.193, HES5, SLCO3A1, and TPM4_8047. In combination, with rPart modelling, the sensitivity and specificity for detecting PCa in blood samples was 78% and 91%, respectively.

Table 8

DMR No.	Marker	AUC	p-value
17	MAX.chr3.193	0.84539	0.1658
12	HES5_0822	0.76068	>0.0001
45	SLCO3A1	0.67209	0.7921
47	TPM4_8047	0.63612	0.227
83	MAX.chr12.48592041-48592162	0.56359	0.037
39	GAS6	0.56073	0.0451
18	MAX.chr3.72788028-72788112	0.53893	0.2411
3	AKR1B1_3644	0.52	0.0165
87	SIM2	0.5101	0.3577
11	HCG4P6	0.5067	0.2852
70	CHST11_2206	0.50524	0.1351
8	GRASP	0.50282	0.2477
44	SERPINB9_3389	0.49417	0.1046
85	PDE4D	0.48252	0.9368
14	ITPRIPL1	0.47	0.7214
16	MAX.chr2.97193166-97193253	0.45301	0.7152

37	FLOT1_1904	0.42456	0.7204
----	------------	---------	--------

Example VII.

Therapeutic decisions for prostate cancer are often guided by Gleason grade, which is subjective and lacks precision. In discovery and early validation, methylated DNA markers (MDMs) were identified with prognostic association (see, Example I). Further experiments were conducted to assess value of novel MDMs in predicting biochemical recurrence using archival tissue from an independent group with > 12 years follow-up after radical prostatectomy (RP).

From 737 men undergoing radical prostatectomy (RP) in 2004, 446 were randomly selected and 155 met quality criteria. Formalin fixed paraffin embedded (FFPE) tissue blocks were utilized. An expert pathologist re-reviewed all specimens in blinded fashion using updated Gleason criteria and marked tumors for macro-dissection. Genomic DNA was purified using the QiaAmp FFPE tissue kit (Qiagen) and quantified by Picogreen fluorescence. Since FFPE DNA can be highly degraded, samples were also tested for amplifiable genome equivalents with a 100bp β -actin amplification assay. DNA was then treated with sodium bisulfite and purified (Zymo Research).

Twenty-three MDMs were chosen to test the samples. The twenty-three MDMs were arrived at by running recursive partitioning analysis (rPART) on the independent tissue validation results which used 73 MDMs. All the patients used in that study (Example I) had outcome data in their clinical record. Specifically, the experiments ran 1000 boot strap sample rPARTs (in-silico) and looked for MDMs which appear most frequently in the modeling trees. These were then ranked by frequency (high to low) and top MDMs were chosen (see, Table 10). MSP assays were run blinded as before against dilutions of universally methylation standards and appropriate negative and positive controls. Raw counts were normalized to overall β -actin counts for each of the samples. Recurrence was defined as PSA > 0.4 ng/mL. Top MDMs were selected by regression partitioning tree models to assign recurrence risk and grouped by quartiles (M1 (lowest) to M4 (highest)). Prognostic values of MDMs and Gleason grade groups (GGG) were assessed and compared based on their concordance with post RP outcomes. The following markers were identified as optimal for predicting recurrence rates WNT3A, LOC100129726, FNBP1, GSDMD, ITPRIPL1, Chr1.61519554, and Chr17.77786040.

Table 10.

Gene	Number Of Models	
WNT3A	582	
LOC100129726	496	
FAM78A	476	
KCNK4	438	
KLF16.RS_FP	415	
CHST11_2206	413	
ARPC1B1906	387	
MAX.chr1.61519554	383	
ABCB1	377	
AGPS_7696	375	
LOC339674	368	
SLCO3A1	368	
LAMA3	367	
MAX.chr17.77786640	366	
AGPS_7497	352	
MAX.chr3.727	351	
GSDMD	339	
GRASP	332	Top 25%
LOC440925	326	
SERPINB9_3389	319	
EIF5A2	318	
MAX.chr10.74079656	315	
MAX.chr15.95128172	311	
SLC4A11	311	
ADD3	309	
ABLIM1	307	
ACOXL	306	
ARPC1B_1967	294	
ANXA2	293	
FLOT1_1586	292	
AKR1B1	287	
HEBP2	287	
S1PR4_9843	287	
FOSL1	285	
MAX.chr3.193776187	282	
HCG4P6	277	
ANKRD35	275	Top 50%
GAS6	270	
LYL1	267	

AGPS_7349	265	
PDE4D	264	
MAX.chr2.97193166	249	
RAI1_7469	237	
FBXO30	232	
STX16	232	
ZNF655_6084	231	
HAPLN3	228	
MAX.chr1.8014263	228	
FLOT1_1904	225	
TPM4_8047	219	
HDAC7_6722	218	
SIM2	216	
SSBP4	215	
RASSF2	214	
MAX.chr16.11327022	210	
GALR3	209	
AKR1B1_3644	208	
ITPRIPL1	204	
Septin9	200	
ADCY9	196	
AXIN1	188	
LRRC4	185	
GNG7_2119	182	
CTBP1	175	
ZNF655_6545	170	
FLJ45983	166	
HES5_1047	162	
NCRNA00092	156	
GSTP1	150	
HES5_0822	148	
FNBP1	132	
BIN2_7908	128	

Example VIII.

Additional experiments were conducted to identify markers capable of distinguishing between 1) PCa Gleason score over 7 versus Gleason score at 6 within prostate tissue, and 2) PCa Gleason score over 6 versus non-cancerous prostate tissue. Such experiments utilized the QuARTs-X (quantitative allele specific real time target and signal assays) (see, e.g., U.S. Patent Application Serial No. 15/335,096).

DNA Extraction

Frozen DNA tissue samples with known clinical information were obtained from the Mayo Clinic repository. DNA was extracted from tissue using DNeasy Blood & Tissue kit from Qiagen per manufacturer protocol. Approximately 100 ng of extracted DNA were

5 carried forward into bisulfite conversion reaction.

Bisulfite conversion and purification of DNA

Reagent Preparation:

Component Abbreviation	Name	Formulation
BIS SLN	Bisulfite Conversion Solution	56.6% Ammonium Bisulfite
DES SLN	Desulphonation Solution	70% Isopropyl alcohol, 0.1 N NaOH
BND BDS	Binding Beads	Maxwell RNA Beads (16 mg/mL), Promega
BND SLN	Binding Solution	7 M Guanidine HCl
CNV WSH	Conversion Wash	10 mM Tris-HCl, 80% Ethanol, 0.01% sodium azide
ELU BUF	Elution Buffer	10mM Tris, 0.1 mM EDTA, pH 8.0

10 Blocking solution:

Number of samples (N)	NaOH (1.6 N) (uL)	BSA (350 ng/ul) (uL)
1	5	5
N		

The following procedure was followed for bisulfite conversion and purification of DNA:

- 15
1. Add 10 μ L of blocking solution to each well in a deep well plate (DWP).
 2. Add 80 μ L of each sample into the DWP.
 3. Carefully mix by pipetting with pipette set to 30-40 μ L to avoid bubbles.

4. Seal and centrifuge the DWP for 1 minute at 3000xg.
5. Incubate at 42 °C for 20 minutes.
6. Add 120 µL of BIS SLN to each well.
7. 8 min cool
- 5 8. Incubate at 65 °C for 75 minutes while mixing during the first 3 minutes.
9. Add 750 µL of BND SLN
10. Pre-mix of silica beads (BND BDS) and add of 50 µL of Silica bead (BND BDS) to the wells of DWP.
11. Mix at 30°C on heater shaker at 1,200 rpm for 30 minutes.
- 10 12. Bead bind on plate magnet for 5 minutes followed by aspiration of solutions to waste.
13. Add 1 mL of wash buffer (CNV WSH) then move the plate to a heater shaker and mix at 1,200 rpm for 3 minutes.
14. Bead bind on plate magnet for 5 minutes followed by aspiration of solutions to waste.
15. Add 0.25 mL of wash buffer (CNV WSH) then move the plate to the heater shaker and
- 15 mix at 1,200 rpm for 3 minutes.
16. Bead bind on the magnet for 2 minutes followed by aspiration of solutions to waste.
17. Add of 0.2 mL of desulphonation buffer (DES SLN) and mix at 1,200 rpm for 7 minutes at 30°C.
18. Bead bind for 2 minutes on the magnet followed by aspiration of solutions to waste.
- 20 19. Add 0.25 mL of wash buffer (CNV WSH) then move the plate to the heater shaker and mix at 1,200 rpm for 3 minutes.
20. Bead bind for 2 minutes on the magnet followed by aspiration of solutions to waste.
21. Add 0.25 mL of wash buffer (CNV WSH) then move the plate to the heater shaker and mix at 1,200 rpm for 3 minutes.
- 25 22. Bead bind for 2 minutes on the magnet followed by aspiration of solutions to waste.
23. Allow plate to dry by moving to heater shaker and incubating at 70°C for 15 minutes while mixing at 1,200 rpm.
24. Add 80 µL of elution buffer (ELU BFR) across all samples in DWP.
25. Incubated at 65 °C for 25 minutes while mixing at 1,200 rpm.
- 30 26. Manually Transfer eluate to 96well plate, seal plate with foil seal, then store at -80°C
27. The recoverable/transferrable volume is about 65 µL.

Multiplex PCR (mPCR) Setup:

1- Prepare a 10X primer mix containing forward and reverse primers for each methylated marker of interest to a final concentration of 750 nM each. Use 10 mM Tris-HCl, pH 8, 0.1 mM EDTA as diluent.

5 2- Prepare 10X mPCR buffer containing 100 mM MOPS, pH 7.5, 75 mM MgCl₂, 0.08% Tween 20, 0.08% IGEPAL CA-630, 2.5 mM dNTPs.

3- Prepare mPCR master mix as follows:

Component	Volume per reaction (μL)
Water	9.62
10X Primer Mix (0.75 uM each)	7.5
mPCR Buffer	7.5
GoTaq (5 units/ul)	0.38
total volume	25.0

4- Thaw DNA and spin plate down.

10 5- Add 25 μL of master mix to a 96 well ABI Veriti plate.

6- Transfer 50 μL of each sample to each well and mix each sample by pipetting up and down several times.

7- Seal Plate with aluminum foil seal

15 8- Place in heated-lid thermal cycler and proceed to cycle using the following profile "QX 12cycle":

Stage	Temp / Time	Number of Cycles
Pre-incubation	95°C / 5 min	1
Amplification 1	95°C / 30 sec	12
	64°C / 60 sec	
Cooling	4°C / hold	1

9- After completion of the incubation, a 1 to 10 dilution of amplicon was performed as follows:

a. Obtain a deep well plate and transfer 180 μL of 10 mM Tris-HCl, pH 8, 0.1 mM EDTA to each well.

- b. Carefully puncture holes into the foil seal of the amplified plate with a 96-well stamp.
- c. Mix the 75 μL amplified sample by repeated pipetting using fresh tips and a 200 μL pipettor set to 50 μL (not generating aerosols) (not using a shaker for mixing)
- d. Add 20 μL of amplified sample to each pre-filled well using fresh tips and a 20 μL pipettor set to 20 μL (not generating aerosols).
- e. Mix the diluted samples by repeated pipetting using fresh tips and a 200 μL pipettor set to 100 μL (not generating aerosols) (not using a shaker for mixing)
- f. Seal the diluted plate with a plastic seal.
- g. Centrifuge the diluted plate at 1,000 rpm for 1 min.
- h. Seal any remaining mPCR product that has not been diluted with a new aluminum foil seal. Place at -80°C .

Manual QuARTS Assay Setup:

- 1- Thaw fish DNA diluent (20 ng/ μL) and use to dilute plasmid calibrators needed in the assay. Use the following table as a dilution guide:

Initial Plasmid Concentration, copies per μL	Final plasmid Concentration, copies per μL	μL of plasmid to add	μL of diluent to add	total volume, μL
1.00E+05	1.00E+04	5	45	50
1.00E+04	1.00E+03	5	45	50
1.00E+03	1.00E+02	5	45	50
1.00E+02	1.00E+01	5	45	50

- 2- Prepare 10X triplex QuARTS oligo mix using the following table for markers A, B, and C:

Oligo	Sequence (5'-3')	Concentration (μM)

Marker A Forward primer	NA	2
Marker A Reverse primer	NA	2
Marker A probe-Arm 1	NA	5
Marker B Forward primer	NA	2
Marker B Reverse primer	NA	2
Marker B probe-Arm 5	NA	5
Marker C Forward primer	NA	2
Marker C Reverse primer	NA	2
Marker C probe-Arm 3	NA	5
A1 HEX FRET	/HEX/ TCT/BHQ- 1/AGCCGGTTTTCCGGCTGAGACCT CGGCG/3C6/ (SEQ ID NO: 235)	5
A5 FAM FRET	/FAM/ TCT/BHQ- 1/AGCCGGTTTTCCGGCTGAGACGT CCGTGG/3C6/ (SEQ ID NO: 236)	5
A3 QUASAR-670 FRET	/Q670/TCT/BHQ- 2/AGCCGGTTTTCCGGCTGAGACTC CGCGTC/3C6/ (SEQ ID NO: 237)	5
dNTP mix		2500

3- Prepare a QuARTS master mix using the following table:

Component	Volume per reaction (μL)
Water	15.5
10X Triplex Oligo Mix	3.0
20X QuARTS Enzyme mix	1.5
total volume	20.0

*20X enzyme mix contains 1 unit/ μ L GoTaq Hot Start polymerase (Promega), 146 ng/ μ L Cleavase 2.0 (Hologic).

- 4- Using a 96 well ABI plates, pipette 20 μ L of QuARTS master mix into each well.
- 5 5- Add 10 μ L of appropriate calibrators or diluted mPCR samples.
- 6- Seal plate with ABI clear plastic seals.
- 7- Centrifuge the plate using 3000rpm for 1 minute.
- 8- Place plate in ABI thermal cycler programmed to run the following thermal protocol “Quarts 5+40” then start the instrument

10

QuARTS Reaction Cycle:				Acquisition
Stage	Temp / Time	Ramp Rate (°C per second)	Number of Cycles	
Pre-incubation	95°C / 3 min	4.4	1	none
Amplification 1	95°C / 20 sec	4.4	5	none
	63°C / 30 sec	2.2		none
	70°C / 30 sec	4.4		none
Amplification 2	95°C / 20 sec	4.4	40	none
	53°C / 1 min	2.2		Yes
	70°C / 30 sec	4.4		none
Cooling	40°C / 30 sec	2.2	1	none

A. Automated QuARTS Setup:

- 1- Thaw 1 tube of fish DNA diluent (20 ng/ μ L), 2 tubes of 1.62X oligo mix, and the prepared calibrator series needed to place on the Hamilton STARlet Deck.
- 15 2- Vortex and centrifuge all reagents prior to loading on the Hamilton STARlet Deck.
- 3- Load the deep well plate containing samples onto the magnet.
- 4- Place full trays of 50 μ L CORE tips on deck as indicated in the diagram below.
- 5- Place at least one full row of 1000 μ L CORE tips on the deck as indicated in the diagram below.
- 20 6- Load an empty ABI 96-well plate onto the STARlet deck with the barcode facing the front of the machine (with A1 well in the back left corner) and as indicated in the diagram below.

- 7- Load reagents in the indicated carrier positions following the on-screen deck layout and software instructions (see second diagram below).
- 8- Load 2 uncapped barcoded empty tubes onto the deck as indicated in the diagram below.
- 5 9- Run the “QuARTSONLYV4.0_BA_20160127” method on the Hamilton.
- 10- Once the method is complete, remove the 96-well QuARTS plate and seal with a clear plastic cover.
- 11- Centrifuge the plate using 3000rpm for 1 minute.
- 12- Place plate in ABI thermal cycler programmed to run the following thermal
- 10 protocol then start the instrument: “Quarts 5+40”.

Results

These experiments determined that

- 1) specific methylated DNA markers (SERPINB9_3479, GRASP_0932,
15 SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075,
PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and
WNT3A_5487) discriminate highly aggressive cancerous prostate tissue (e.g., Gleason score
at or above 7.0 (e.g., 7, 8, 9, 10) from less-aggressive cancerous prostate tissue (e.g., Gleason
score below 7 (e.g., 6), and
- 20 2) specific methylated DNA markers (SERPINB9_3479, FLOT1_1665,
HCG4P6_4618, CHST11_2206, MAX.chr12.485, GRASP_0932, GAS6_6425,
MAX.chr3.193, MAX.chr2.971_3164, MAX.chr3.727_8028, HES5_0840, TPM4_8037,
SLCO3A1_6187, ITPRIPL1_1244, AKR1B1_3644, RASGRF2_6325, ZNF655_6075,
PAMR1_7364, ST6GALNAC2_1113, CCNJL_9070, KCNB2_9128, IGFBP7_6412, and
25 WNT3A_5487) discriminate cancerous prostate tissue (e.g., Gleason score at or above 6.0
(e.g., 6, 7, 8, 9, 10) from non-cancerous prostate tissue.

Table 11 shows the % methylation for normal tissue, % methylation for prostate tissue having Gleason Score of 6, and % methylation for prostate tissue having Gleason Score between 7-10 (the oligo sequences are provided in Table 12; the DMR information is
30 provided in Table 13).

Table 11.

Marker	% methylation Normal Tissue (28 samples)	% methylation Gleason Score 6 (24 samples)	% methylation Gleason Score 7-10 (42 samples)
SERPINB9_3479	0.42	12.77	19.45
FLOT1_1665	5.82	42.61	41.91
HCG4P6_4618	3.64	31.43	38.48
CHST11_2206	0.38	16.17	21.95
MAX.chr12.485	3.46	28.85	32.24
GRASP_0932	0.59	28.31	45.31
GAS6_6425	0.34	21.72	20.92
MAX.chr3.193	1.61	27.69	33.89
MAX.chr2.971_3164	2.08	19.89	24.66
MAX.chr3.727_8028	0.72	19.76	27.73
HES5_0840	0.76	13.81	15.08
TPM4_8037	0.16	3.19	2.90
SLCO3A1_6187	0.10	3.77	7.18
ITPRIPL1_1244	0.18	8.77	17.76
AKR1B1_3644	0.10	5.82	8.32
JSRP1	0.09	2.16	7.61
RASGRF2_6325	0.02	0.61	3.07
ZNF655_6075	0.35	20.51	49.64
PAMR1_7364	0.03	0.40	1.79
ST6GALNAC2_1113	0.36	13.76	19.78
CCNJL_9070	0.15	3.33	7.85
KCNB2_9128	0.19	8.48	17.22
IGFBP7_6412	0.05	1.06	3.16

WNT3A_5487	11.52	19.60	37.38
------------	-------	-------	-------

Notes:

1. Gleason Score of 7-10 comprising
 - a. 21 samples of Gleason Score 7
 - b. 6 samples of Gleason Score 8
 - c. 11 samples of Gleason Score 9
 - d. 4 samples of Gleason Score 10
 2. % methylation is [(number of strands marker) ÷ (number of strands actin)] x 100
- Using a logistic regression analysis fit of % methylation relative to ACTB on this data, using a 100% cutoff, prediction of cancer from normal with 98.5% sensitivity using two markers, FLOT1 and MAX.Chr3.193, with an AUC = 0.99, was permitted. For the 6 vs 6+ prediction, a logistic regression analysis fit of % methylation of 9 markers relative to ACTB of the data, predicted the Gleason 6+ with 92.8% sensitivity at a specificity of 91.7% (AUC = 0.96). The markers were: GRASP, GAS6, MAX.chr3.193, MAX.chr2.971, TPM4, ITPRIPL2, AKR1B1, ZNF655, WNT3A.

Table 12.

MARKER	OLIGO TYPE	OLIGO NAME	SEQUENCE
JSRP1	Primer	JSRP1_FP	GGGTCGTTAGGAGTGTTCG (SEQ ID NO: 207)
	Primer	JSRP1_RP	CCTCTCTAAAAACCGCTCAAC (SEQ ID NO: 208)
	Probe	JSRP1_Pb_A5	CCACGGACG CTCGTAAACGCC/3C6/ (SEQ ID NO: 209)
ZNF655_6075	Primer	ZNF655_6075_F P	AAGACGTGGAAAAGTTGCG (SEQ ID NO: 210)
	Primer	ZNF655_6075_R P	CCGCGCGTCCATAATTC (SEQ ID NO: 211)
	Probe	ZNF655_6075_P b_A1	CCACGGACG CGAAACTCGTT/3C6/ (SEQ ID NO: 212)
KCNB2_9128	Primer	KCNB2_9128_F P	GTAGGAGTGGTTGGCGC (SEQ ID NO: 213)

	Primer	KCNB2_9128_R P	CCCACACCTCGACGAAAT (SEQ ID NO: 214)
	Probe	KCNB2_9128_P b_A5	CCACGGACG CGCGGAAGTTGA/3C6/ (SEQ ID NO: 215)
ST6GALNA C2_1113	Primer	ST6GALNAC2_1 113_FP	GGAGGAGAACGCGGATG (SEQ ID NO: 216)
	Primer	ST6GALNAC2_1 113_RP	GCGATCCGCGAAAAACG (SEQ ID NO: 217)
	Probe	ST6GALNAC2_1 113_Pb_A1	CCACGGACG GAACGCCCCGAAA/3C6/ (SEQ ID NO: 218)
MAX.chr3.7 27_8028	Primer	MAX.chr3.727_8 028_FP	GTGGTTTTATTTCTGTTTCGTTTCG (SEQ ID NO: 177)
	Primer	MAX.chr3.727_8 028_RP	AAACTAACGAATATAACTCCGTCGA (SEQ ID NO: 178) or CTAACTAACTAACGAATATAACTC CGTC (SEQ ID NO: 219)
	Probe	MAX.chr3.727_8 028_Pb_A1	CGCCGAGG GCGACGTTTCGAG/3C6/ (SEQ ID NO: 179)
RASGRF2_ 6325	Primer	RASGRF2_6325 _FP	GTTAGGGCGGAGAGCGT (SEQ ID NO: 220)
	Primer	RASGRF2_6325 _RP	CGCGCGATAACAAAAACG (SEQ ID NO: 221)
	Probe	RASGRF2_6325 _Pb_A5	CGCCGAGG GCGAACTAAAAC/3C6/ (SEQ ID NO: 222)
PAMR1_73 64	Primer	PAMR1_7364_F P	ACGTTTGGAGATTCGCGG (SEQ ID NO: 223)
	Primer	PAMR1_7364_R P	CCCCGCAACTTCCTT (SEQ ID NO: 224)
	Probe	PAMR1_7364_P b_A5	CGCCGAGG GACGGCGGTTGT/3C6/ (SEQ ID NO: 225)
IGFBP7_64 12	Primer	IGFBP7_6412_F P	GGGTCGTAGGTGTTTCGAA (SEQ ID NO: 226)
	Primer	IGFBP7_6412_R P	GCGCCCTACTCCTCGAC (SEQ ID NO: 227)
	Probe	IGFBP7_6412_P b_A5	CGCCGAGG CGCCGCTAACT/3C6/ (SEQ ID NO: 228)
CCNJL_907 0	Primer	CCNJL_9070_F P	GGTATCGTAGTTTTTCGCGGA (SEQ ID NO: 229)
	Primer	CCNJL_9070_R P	CTCCTACGCCGCTCAA

			(SEQ ID NO: 230)
	Probe	CCNJL_9070_Pb_A5	CGCCGAGG ATTAGAGGCGAT/3C6/ (SEQ ID NO: 231)
WNT3A_5487	Primer	WNT3A_5487_F_P	GTGTAAATGCGCGGGC (SEQ ID NO: 232)
	Primer	WNT3A_5487_R_P	CGCTTTAATTCAACACCGCG (SEQ ID NO: 233)
	Probe	WNT3A_5487_Pb_A5	CGCCGAGG CGGTTTATACGT/3C6/ (SEQ ID NO: 234)
FLOT1_1665	Primer	FLOT1_1665_FP	GTAGTGTTTTGAGTTTAAGTTGTTT CG (SEQ ID NO: 150)
	Primer	FLOT1_1665_R_P	AACACCAAACGTACCTAATACG (SEQ ID NO: 151)
	Probe	FLOT1_1665_Pb_A1	CGCCGAGG GCGAAAACGTAT/3C6/ (SEQ ID NO: 152)
HCG4P6_4618	Primer	HCG4P6_4618_FP	CGGAGTATACGGAATATTAGGTTCCG (SEQ ID NO: 153)
	Primer	HCG4P6_4618_RP	AATAATAACGACGCGACATAAACA (SEQ ID NO: 154)
	Probe	HCG4P6_4618_Pb_A5	CCACGGACG GCGTATAGATTG/3C6/ (SEQ ID NO: 155)
SERPINB9_3479	Primer	SERPINB9_3479_FP	TTTATTTTAGTCGTGCGCGG (SEQ ID NO: 147)
	Primer	SERPINB9_3479_RP	ACGCGACACGAAAACGAC (SEQ ID NO: 148)
	Probe	SERPINB9_3479_Pb_A5	CCACGGACG CGACAACGCTAA/3C6/ (SEQ ID NO: 149)
CHST11_2206	Primer	CHST11_2206_F_P	GCGTTCGAGGGCGTTTT (SEQ ID NO: 156)
	Primer	CHST11_2206_RP	AATCTAACTAAAAAAAAAACGAA CTCGT (SEQ ID NO: 157)
	Probe	CHST11_2206_Pb_A1	CGCCGAGG TCGCTTCCTAAA/3C6/ (SEQ ID NO: 158)
MAX.chr12.485	Primer	MAX.chr12.485_FP	GAGGAAATAGGCGTGGTTCG (SEQ ID NO: 159)
	Primer	MAX.chr12.485_RP	AAAACCGAAAAAACTAATTCGTCA (SEQ ID NO: 160)
	Probe	MAX.chr12.485_Pb_A5	CCACGGACG GCGTTGACGGTT/3C6/ (SEQ ID NO: 161)
GRASP_0932	Primer	GRASP_0932_F_P	CGCGCGGTATAGTTCGG (SEQ ID NO: 162)
	Primer	GRASP_0932_R_P	ACCCCAACGAACGAAC (SEQ ID NO: 163)

	Probe	GRASP_0932_P b_A1	CGCCGAGG CTATACGCGACG/3C6/ (SEQ ID NO: 164)
GAS6_6425	Primer	GAS6_6425_FP	GATTTTGAGGTTAGGTTTCGTCG (SEQ ID NO: 165)
	Primer	GAS6_6425_RP	GAAAAACAACGCTACTACCGC (SEQ ID NO: 166)
	Probe	GAS6_6425_Pb_ A5	CCACGGACG GCGCGGAGTTGG/3C6/ (SEQ ID NO: 167)
MAX.chr3.1 93	Primer	MAX.chr3.193_F P	TTAAAGGTTTAGTTAAAGATGGAAT CGT (SEQ ID NO: 174)
	Primer	MAX.chr3.193_R P	CTCGCGACGAAAAAACCC (SEQ ID NO: 175)
	Probe	MAX.chr3.193_P b_A1	CGCCGAGG CGAACTCCCAAC/3C6/ (SEQ ID NO: 176)
MAX.chr2.9 71	Primer	MAX.chr2.971_F P	GCGAGGTTGCGGTTTTG (SEQ ID NO: 198)
	Primer	MAX.chr2.971_R P	CGAAAAATAATTGATTAACCAAA ACGC (SEQ ID NO: 199)
	Probe	MAX.chr2.971_P b_A5	CCACGGACG CCGATATTTCCC/3C6/ (SEQ ID NO: 200)
HES5_0840	Primer	HES5_0840_FP	GCGAGAGGAGTAGGTTCG (SEQ ID NO: 180)
	Primer	HES5_0840_RP	AACCTACGAACGCGCGA (SEQ ID NO: 181)
	Probe	HES5_0840_Pb_ A5	CCACGGACG ACCCGACGACCA/3C6/ (SEQ ID NO: 182)
TPM4_8037	Primer	TPM4_8037_FP	GGAGAAAGGCGGGCG GA (SEQ ID NO: 189)
	Primer	TPM4_8037_RP	CGTTCCCAAAAACGCGA (SEQ ID NO: 190)
	Probe	TPM4_8037_Pb_ A1	CGCCGAGG ACGACGACGTAT/3C6/ (SEQ ID NO: 191)
SLCO3A1_6 187	Primer	SLCO3A1_6187 _FP	CGATTTTGTAGCGTTTGGATATCG (SEQ ID NO: 171)
	Primer	SLCO3A1_6187 _RP	TCGAAAACCCACGAACC (SEQ ID NO: 172)
	Probe	SLCO3A1_6187 _Pb_A5	CCACGGACG CGTTAATACCCC/3C6/ SEQ ID NO: 173)
ITPRIPL1_1 244	Primer	ITPRIPL1_1244_ FP	GTAGTGGCGGTTAGGTTCG (SEQ ID NO: 195)
	Primer	ITPRIPL1_1244_ RP	TCACCGTCAATATTAATAAAAACAC GA (SEQ ID NO: 196)
	Probe	ITPRIPL1_1244_ Pb_A1	CGCCGAGG GCGTTGTTTTTT/3C6/ (SEQ ID NO: 197)

AKR1B1_3644	Primer	AKR1B1_3644_F P	GTTCGTTTTATATTTTCGCGCG (SEQ ID NO: 201)
	Primer	AKR1B1_3644_ RP	CCGAACGTCCGCGAAAC (SEQ ID NO: 202)
	Probe	AKR1B1_3644_ Pb_A1	CGCCGAGG CGAACTACTCAA/3C6/ (SEQ ID NO: 203)

Table 13.

DMR No.	Gene Annotation	Chromosome No.	Region on Chromosome (starting base-ending base)
121	WNT3A_5487	1	228225487-228225590
122	SERPINB9_3479	6	2903389-2903479
123	FLOT1_1665	6	30711586-30711665
124	HCG4P6_4618	6	29894618-29894693
125	GRASP_0932	12	52400932-52401020
126	GAS6_6425	13	114566425-114566518
127	SLCO3A1_6187	15	92396091-92396187
128	MAX.chr3.727_8028	3	72788028-72788112
129	HES5_0840	1	2460840-2460903
130	TPM4_8037	19	16188037-16188154
131	ITPRIPL1_1244	2	96991244-96991312
132	MAX.chr2.971_3164	2	97193164-97193252
133	AKR1B1_3644	7	134143644-134143721
134	RASGRF2_6325	5	80256325-80256390
135	PAMR1_7364	11	35547364-35547423
136	ZNF655_6075	7	99156075-99156154
137	CCNJL_9070	5	159739070-159739148
138	ST6GALNAC2_1113	17	74581113-74581238
139	IGFBP7_6412	4	57976412-57976506
140	KCNB2_9128	8	73449128-73449208
17	MAX.chr3.193	3	193776187-193776257
70	CHST11_2206	12	104852206-104852307
83	MAX.chr12.485	12	48592041-48592162

5 All publications and patents mentioned in the above specification are herein incorporated by reference in their entirety for all purposes. Various modifications and variations of the described compositions, methods, and uses of the technology will be

apparent to those skilled in the art without departing from the scope and spirit of the technology as described. Although the technology has been described in connection with specific exemplary embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in pharmacology, biochemistry, medical science, or related fields are intended to be within the scope of the following claims.

It is an object of the present intention to overcome or ameliorate at least one of the disadvantages of the prior art, or to provide a useful alternative.

The term 'comprise' and variants of the term such as 'comprises' or 'comprising' are used herein to denote the inclusion of a stated integer or stated integers but not to exclude any other integer or any other integers, unless in the context or usage an exclusive interpretation of the term is required.

Any reference to publications cited in this specification is not an admission that the disclosures constitute common general knowledge.

A definition of the specific embodiment of the invention claimed herein follows.

In a broad format, the invention provides a method for detecting prostate cancer, comprising:

- a) extracting genomic DNA from a biological sample of a human individual suspected of having or having prostate cancer;
- b) treating the extracted genomic DNA with bisulfite;
- c) amplifying the bisulfite-treated genomic DNA with primers consisting of a pair of primers specific for one or more CpG sites in the gene SERPINB9; and
- d) measuring the methylation level of the one or more CpG sites in SERPINB9 using methylation-specific PCR, quantitative methylation-specific PCR, methylation sensitive DNA restriction enzyme analysis or bisulfite genomic sequencing PCR.

CLAIMS**WE CLAIM:**

1. A method for detecting prostate cancer, comprising:
 - a) extracting genomic DNA from a biological sample of a human individual suspected of having or having prostate cancer;
 - b) treating the extracted genomic DNA with bisulfite;
 - c) amplifying the bisulfite-treated genomic DNA with primers consisting of a pair of primers specific for one or more CpG sites in the gene SERPINB9; and
 - d) measuring the methylation level of the one or more CpG sites in SERPINB9 using methylation-specific PCR, quantitative methylation-specific PCR, methylation sensitive DNA restriction enzyme analysis or bisulfite genomic sequencing PCR.

2. The method of claim 1, wherein the amplifying step (c) further comprises amplifying the bisulfite-treated genomic DNA with primers consisting of a pair of primers specific for one or more CpG sites in a gene selected from FLOT1, HCG4P6, CHST11, MAX.chr12.485, GRASP, GAS6, MAX.chr3.193, MAX.chr2.971, MAX.chr3.727, HES5, TPM4, SLCO3A1, ITPRIPL1, AKR1B1, RASGRF2, ZNF655, PAMR1, ST6GALNAC2, CCNJL, KCNB2, IGFBP7, and WNT3A.

3. The method of claim 1 or claim 2, wherein the measuring step (d) further comprises measuring the methylation level of one or more CpG sites in a gene selected from FLOT1, HCG4P6, CHST11, MAX.chr12.485, GRASP, GAS6, MAX.chr3.193, MAX.chr2.971, MAX.chr3.727, HES5, TPM4, SLCO3A1, ITPRIPL1, AKR1B1, RASGRF2, ZNF655, PAMR1, ST6GALNAC2, CCNJL, KCNB2, IGFBP7, and WNT3A.

4. The method of any one of claims 1-3, wherein the sample is a prostate tissue sample.

5. The method of any one of claims 1-4, wherein:
 - the set of primers for MAX. chr3.727 consist of SEQ ID NOs: 177 and 178, or SEQ ID NOs: 177 and 219;
 - the set of primers for RASGRF2 consist of SEQ ID NOs: 220 and 221;
 - the set of primers for ZNF655 consist of SEQ ID NOs: 210 and 211;

the set of primers for PAMR1 consist of SEQ ID NOs: 223 and 224;
the set of primers for ST6GALNAC2 consist of SEQ ID NOs: 216 and 217;
the set of primers for CCNJL consist of SEQ ID NOs: 229 and 230;
the set of primers for KCNB2 consist of SEQ ID NOs: 213 and 214;
the set of primers for IGFBP7 consist of SEQ ID NOs: 226 and 227;
the set of primers for WNT3A consist of SEQ ID NOs: 232 and 233;
the set of primers for SERPINB9 consist of SEQ ID NOs: 147 and 148, or SEQ ID NOs: 129 and 130;

the set of primers for FLOT1 consist of SEQ ID NOs: 150 and 151;
the set of primers for HCG4P6 consist of SEQ ID NOs: 153 and 154;
the set of primers for CHST11 consist of SEQ ID NOs: 156 and 157;
the set of primers for MAX.chr12.485 consist of SEQ ID NOs: 159 and 160;
the set of primers for GRASP consist of SEQ ID NOs: 162 and 163;
the set of primers for GAS6 consist of SEQ ID NOs: 165 and 166;
the set of primers for MAX.chr3.193 consist of SEQ ID NOs: 174 and 175;
the set of primers for MAX.chr2.971 consist of SEQ ID NOs: 198 and 199;
the set of primers for HES5 consist of SEQ ID NOs: 180 and 181;
the set of primers for TPM4 consist of SEQ ID NOs: 189 and 190;
the set of primers for SLCO3A1 consist of SEQ ID NOs: 171 and 172;
the set of primers for ITPRIPL1 consist of SEQ ID NOs: 195 and 196; and
the set of primers for AKR1B1 consist of SEQ ID NOs: 201 and 202.

6. The method of any one of claims 1-5, wherein the amplifying step (c) further comprises amplifying bisulfite-treated genomic DNA with primers consisting of a pair of primers specific for one or more CpG sites in a gene selected from ACOXL, ANXA2, FLJ45983, HAPLN3, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.72788028-72788112, RAI1, RASSF2, and SLC4A11.

7. The method of any one of claims 1-6, wherein the measuring step (d) further comprises measuring the methylation level of one or more CpG sites in a gene selected from ACOXL, ANXA2, FLJ45983, HAPLN3, KCNK4, MAX.chr1.61519554-61519667, MAX.chr2.97193166-97193253, MAX.chr3.72788028-72788112, RAI1, RASSF2, and SLC4A11.

8. The method of claim 6 or claim 7, wherein:
- the set of primers for ACOXL consist of SEQ ID NOs: 93 and 94;
 - the set of primers for ANXA2 consist of SEQ ID NOs: 89 and 90;
 - the set of primers for FLJ45983 consist of SEQ ID NOs: 31 and 32;
 - the set of primers for HAPLN3 consist of SEQ ID NOs: 37 and 38;
 - the set of primers for KCNK4 consist of SEQ ID NOs: 125 and 126;
 - the set of primers for MAX.chr1.61519554-61519667 consist of SEQ ID NOs: 91 and 92;
 - the set of primers for MAX.chr2.97193166-97193253 consist of SEQ ID NOs: 49 and 50;
 - the set of primers for MAX.chr3.72788028-72788112 consist of SEQ ID NOs: 53 and 54;
 - the set of primers for RAI1 consist of SEQ ID NOs: 55 and 56;
 - the set of primers for RASSF2 consist of SEQ ID NOs: 57 and 58; and
 - the set of primers for SLC4A11 consist of SEQ ID NOs: 59 and 60.

SEQUENCE LISTING

<110> Mayo Foundation for Medical Education and Research
EXACT SCIENCES DEVELOPMENT COMPANY, LLC

<120> DETECTING PROSTATE CANCER

<130> EXCTM-35073/WO-1/ORD

<150> US 62/464,800

<151> 2017-02-28

<160> 237

<170> PatentIn version 3.5

<210> 1

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 1

c g a g t t g t a t

a g t t a g a a g a

g g a c g t

26

<210> 2

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 2

a a c c g a a a a a

a c c t a a t t c g

a a a c g

25

<210> 3

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 3

g g g g t a t t a g

t t t t t t a t g a

g a t t g c g t

28

<210> 4

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 4

t a a a a a a a t c

c a c c t a t c g c

t c g a a

25

<210> 5

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 5

t t g t t a a t t t

t t t g g g g g t c

g t c g t

25

<210> 6

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 6

g a a a a c c c g c
23

a c t t c c t c c t

c g a

<210> 7
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 7
t a t a g g a t t t
25

t a a g t t c g c g

a a c g t

<210> 8
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 8
t c a a c c t a c t
28

t c c t t c c t a t

a t a a c g a t

<210> 9
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 9
g g a g t t a g a a
26

g t t t t t t g g g

a g g g g c

<210> 10
<211> 25

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 10
a a a c a c g t a a a c c c t t c t a c c c g a c
25

<210> 11
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 11
g c g t g a t t g a t g g g t g t a t t a c g t
24

<210> 12
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 12
a t a a a c t t c c g a t c c c t a c a a c g a a
25

<210> 13
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 13
t a g t t t c g g g g t t g c g g t g a t t t a c
25

<210> 14
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 14
a a a a t c c g a a a c a a a c a t t c a c g c c
25

<210> 15
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 15
t g t t t c g t t t a g t t a t g a a g t a t a t c g g
28

<210> 16
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 16
a c t t a a t c g c g c a a c a a c c g c a
22

<210> 17

<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 17
c g g t t a t a t t a t a a g a a a g g a a g t t t t c g t
31

<210> 18
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 18
g a a a a c c c a a a c t a c a c a c c g c t
23

<210> 19
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 19
t t g g t t t t t a t t a a g t t a t g t g a c g a
26

<210> 20
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 20
t a a a a t t c c a a a a a c g a t a a c g c t
25

<210> 21
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 21
t g a g g t t a a g g g a t a g t t t t c g c g g
25

<210> 22
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 22
a a c t a c a a c t t c a a a t a c t c c g c a c g c t
28

<210> 23
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 23
g g t t t t g c g t t t t g c g t t c g c
21

<210> 24
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 24
c c a t t a t a t a a a c t t c c c c t t c g c c
25

<210> 25
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 25
t t t c g g g c g t t t t a g g t t c g t t t c
24

<210> 26
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 26
g a c t c a a c g a t a c t c c c a c c g c c
23

<210> 27
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 27

g g g g t t c g t t
25

t t a t a t t t t t

c g c g c

<210> 28

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 28

c a a a a t c a c t
24

c g a a a t c c c t

c g c c

<210> 29

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 29

g g g a g g t a g t
28

t a g t t t a c g g

t a a t a c g a

<210> 30

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 30

c t a a a c c a c c
23

a a c g a a c c c c

g a a

<210> 31
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 31
t a g t c g a g g t t a t g g a g g t g a c g g c
25

<210> 32
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 32
a c t a c c c g t t a a a c a c g a c g a a
22

<210> 33
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 33
t g t t t t c g g a t a c g g c g a g c
20

<210> 34
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 34

a c g a a c g a a c

t a t a c g c g a c

g c t

23

<210> 35

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 35

t c g t g a t t t a

g t a t t g g g g c

g g a g c

25

<210> 36

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 36

g a a a c t c c a a

c g a a a a c c t c

g c g a c

25

<210> 37

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 37

a a g c g g t a a g

g g a g g a a t t c

g g t t c

25

<210> 38
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 38
g a c c c c c g a a a a c t c t a a c c g t c g
24

<210> 39
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 39
g g a t c g g a g t a t t g g g a t c g g a g t a t a c
28

<210> 40
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 40
a a c t c t a a t a a t a a t a a c g a c g c g a c
26

<210> 41
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 41

g t g a t t c g g c
20

g g g a t t t g c g

<210> 42

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 42

g a c g a a a a a a
27

c g c t t c c c t a

c a a a c g a

<210> 43

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 43

a a g a g t t t g t
26

a t t a g g a t t a

t a g c g a

<210> 44

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 44

c g t a a a a c g t
25

c a a a a a c t a c

a c g a c

<210> 45
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 45
g g g a t t t a g g g t t a g g t t a t t t a t c g t
27

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 46
a c c g c g c t a t c t c t t t a a a a t c g t a
25

<210> 47
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 47
t a a t t t c g c g a g g t a g g c g a c g g
23

<210> 48
<211> 26
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 48

c a a t a c t c t t
26

a t a t a t t a a c

g c c g c t

<210> 49

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 49

g g t t t t a g g g
22

a a a t a t c g g c

g t

<210> 50

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 50

a a c t c a a a c c
26

g a a a a a t a a

t t c g a t

<210> 51

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 51

a a a g g t t t a g
26

t t a a a g a t g g

a a t c g t

<210> 52
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 52
c t c g c g a c g a a a a a a a c c c g a a
22

<210> 53
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 53
a g g a t t c g a c g g a g t t a t a t t c g t
24

<210> 54
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 54
t a a c a t a a c c a c c c a a c t c t c c c c g a a
27

<210> 55
<211> 34
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 55

t t a t a g t a g t t c g t c g a a a a t a t a a a g t t t c g t c
34

<210> 56

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 56

c c g a a a a a c c c a a a a a a a a c c c g
23

<210> 57

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 57

g t c g t t t a c g g t a t t t g t t t c g t t c
25

<210> 58

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 58

a t c g c t a a a a c c t c a a c c t a a c g t c
25

<210> 59
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 59
t t c g g t t g t t t g t g t t t c g t t g t c g
25

<210> 60
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 60
c c g a a t c t a a a a a c g c t t t c c t c t c g t a
28

<210> 61
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 61
g t a g c g t c g g c g t a g a g c g t a g a t c
25

<210> 62
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 62

a t a a c g a a t c
24

c c c g c g a a a a

t t c g

<210> 63

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 63

g g g g t a g a g a
28

a t g t g a a g t t

t t a g a c g t

<210> 64

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 64

a c c g c g a c g a
20

c t t a a c g a c g

<210> 65

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 65

t t t t t t a t t c
26

g c g t t t a g c g

g t t t c g

<210> 66
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 66
c c g c c a t a a c t a c c g c c t t a t a c t a c c g
28

<210> 67
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 67
t t a a t g a g t g t a a a g c g c g g a g a g t c g
27

<210> 68
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 68
c g a a a a t a a c c g t a a a c g c t a c c g t
25

<210> 69
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 69

t g t a g t a g g a
25

t a c g t t g a g t

c g c g g

<210> 70

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 70

g c g a t a a a a c
21

t c c a c g c c g t

t

<210> 71

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 71

g t a g t a t t t t
25

a t t t t t a t c g

g g c g a

<210> 72

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 72

a a a a a a t c a c
25

a c c t c c t t c a

a c g c t

<210> 73
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 73
a a t a c g c g a c g g t t t c g t t t a t t g c
25

<210> 74
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 74
c a t a a c g t a a a t c c a c t t c c g a c g a c
26

<210> 75
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 75
c g t t t a g g g g g t t t c g a g c g t a g c
24

<210> 76
<211> 27
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 76

c c c t a a t t c c
27

a t c c t a a a c g

a a t c g a c

<210> 77

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 77

t t g c g a a a a c
29

g a g t t t t c g a

a t t a t g g a c

<210> 78

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 78

c c c c g a a t a t
26

a a a t a a c g a c

c c c g a a

<210> 79

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 79

t t g g t t t a t t
25

t a t t t c g c g g

a t c g a

<210> 80
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 80
a a a a c a c g a t c g c c g a c t c c t a a c g
25

<210> 81
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 81
a g g g a g t t t t t c g t t g g t t g t c g a c
25

<210> 82
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 82
a t a c t a c g a a c t c c g c g c t c a c a c g
25

<210> 83
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 83

a g g g a g t t t t
25

t c g t t g g t t g

t c g a c

<210> 84

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 84

a t a c t a c g a a
25

c t c c g c g c t c

a c a c g

<210> 85

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 85

t t t t t t t a g t
25

t t a g a t t t c g

g a c g t

<210> 86

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 86

t a a a a a t a a a
25

c c c c a t t c c t

c c g a t

<210> 87
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 87
c g g t g a g t a t t t c g c g g t t t t t c g t
25

<210> 88
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 88
a a a t c g c c t c t c c c g a a c g c
20

<210> 89
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 89
g g g t t t a a a g t t a t g g g t t t t a t t t c g t
28

<210> 90
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 90

t a a a a c a t c t
25

c t a c g c g a c c

t c g t a

<210> 91

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 91

g t c g c g t t t t
26

t t t a t t t g t c

g t t c g t

<210> 92

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 92

a t a a a a a t c g
26

a t t c t a c c g c

g t c g c c

<210> 93

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 93

a g t t a a g t t t
25

t a a c g g g t g t

g g c g g

<210> 94
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 94
a a a c g t c g a t a a a a c g a a c g t c g t a
25

<210> 95
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 95
t g a g a t t t c g g t t a a a g g a a g g g g t c
26

<210> 96
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 96
a c c t a c c t a t a c g c c t c c a a a c g a t
25

<210> 97
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 97

a a t t t c g t a g
25

g a a g t a g t c g

g t c g t

<210> 98

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 98

t a c c c t c c a a
25

a t t a c c c a a a

c c g c t

<210> 99

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 99

t t g t t t t t t g
21

a g t t c g c g g g

c

<210> 100

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 100

a c c a a t a c g a
24

t t c t c c c t c c

c g a t

<210> 101
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 101
t c g c g t t g a c g g t t t g t g a c g
21

<210> 102
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 102
a a c c c c g a t c c g a a a a a c c g a a
22

<210> 103
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 103
t t c g a g g t t t t t c g g g g c g a c
21

<210> 104
<211> 26
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 104

c c c t c c t a t c
26

c g t t a a c t t t

c g c g t a

<210> 105

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 105

g g t c g t a g t t
22

c g g g a a g t t c

g g

<210> 106

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 106

a t t c g a c g a a
20

a a c a c c g c g c

<210> 107

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 107

c g g g a a a t t t
27

t c g g a a a a t a

t a t a c g t

<210> 108
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 108
t a c g a a a t t c c a a c a a a a a c c g a a
25

<210> 109
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 109
c g g g c g g t t t a t a c g t t t t t c g c
23

<210> 110
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 110
a a a a a a a t c c c a t t c a a a c g c t
24

<210> 111
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 111

t g g t g g a t c g
25

c g a t t t t c g t

a a g a c

<210> 112

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 112

c g c c g a a a a c
22

c a a a t t a t c g

c g

<210> 113

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 113

g g g t g g a t t t
26

t c g g g c g t t a

t a a a t c

<210> 114

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 114

c a a a a c g a c t
20

c c c c g c c g a a

<210> 115
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 115
g t t t t t a t t t g g a t g g t g t g g t c g t
25

<210> 116
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 116
c c c g a a c t c t a c t a c a a a t t a a t c g a a
27

<210> 117
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 117
t a g t t t a g t a g a g g g a g g g t c g c g g
25

<210> 118
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 118

c g c g a a a a a
23

c c g a a a a t c c

g t t

<210> 119

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 119

g t t a t t a g g a
25

t t t g g t a g a a

g a c g a

<210> 120

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 120

a a c t a c c a a t
22

c g a a a a a c c g

a a

<210> 121

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 121

t g t t t c g g a a
29

g t t t t a g t t g

g g g a t a c g t

<210> 122
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 122
a a c a c c a a a c g t a c c t a a t a c g c g a a
26

<210> 123
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 123
t t t t t t t a t t t t t t a t t t t t t t c g t c g t
28

<210> 124
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 124
g a c t t c t a c t t a c t t c c c t a a c c g t t
26

<210> 125
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 125

t t a g t t a g g a a g t a g g g c g a g g c g a
25

<210> 126

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 126

a a a a c c g a a c a a c g c a a a a c g a a
23

<210> 127

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 127

t t t c g a c g a g t a g g a t t g a a g a a g g a a c g
29

<210> 128

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 128

g c g a a t c t a t c t a c c g a a a c g c g c t
25

<210> 129
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 129
g g t t t t a t t a t t c g g t g g t a g t c g g
25

<210> 130
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 130
a c a c g a a a a c g a c g a c a a c g c t
22

<210> 131
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 131
g t t a g g a a g g g a g g g a c g t t t c g g
24

<210> 132
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 132

c g a a a a c t a
25

c g a a c a c g t a

t c g a c

<210> 133

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 133

t a c g t c g t t c
34

g t t a t t t a g a

t t t a t a a t t t

t g t c

<210> 134

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 134

c c a a a c c c t a
34

c t a c t a c t a c

t a c t a c t a a c

t a c g

<210> 135

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 135

a t c g t t t t a t
25

c g t a g a a g t c

g g c g a

<210> 136
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 136
t a c g a c c t a a a c t a a a t c c c c c g c a
25

<210> 137
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 137
g c g a t t t t g t a g c g t t t g g a t a t c g a
26

<210> 138
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 138
g t c g a a a a c c c c a c g a a c c g t t
22

<210> 139
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 139

g g t t g t a a g g
24

g g g t t t g g g t

a c g c

<210> 140

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 140

a t t t c g a a a a
25

a a a c g c c c g a

t a c g a

<210> 141

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 141

g t t t t t t c g t
25

c g g t t a a t t t

a g c g t

<210> 142

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 142

a a a a a t a a c g
25

a a t c a c a a c a

c c g t c

<210> 143
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 143
g a t t t c g g g c g t a g a t a g g g a c g t
24

<210> 144
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 144
a c a a c c g a a c t a c a a a t a c c t c g a a
25

<210> 145
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 145
g t t c g t t t a g a a g g t t t c g c g t c g t a t a c
29

<210> 146
<211> 26
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 146

a c c t t t c c c g
26

a a a c c t a a a a

c t t c c g

<210> 147

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 147

t t t a t t t t a g
20

t c g t g c g c g g

<210> 148

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 148

a c g c g a c a c g
18

a a a a c g a c

<210> 149

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 149

c c a c g g a c g c
22

g a c a a c g c t a

a c

<210> 150
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 150
g t a g t g t t t t g a g t t t a a g t t g t t t c g
27

<210> 151
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 151
a a c a c c a a a c g t a c c t a a t a c g
22

<210> 152
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 152
c g c c g a g g g c g a a a a c g t a t c
21

<210> 153
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 153

c g g a g t a t a c
25

g g a a t a t t a g

g t t c g

<210> 154

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 154

a a t a a t a a c g
24

a c g c g a c a t a

a a c a

<210> 155

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 155

c c a c g g a c g g
22

c g t a t a g a t t

g c

<210> 156

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 156

g c g t t c g a g g
17

g c g t t t t

<210> 157
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 157
a a t c t a a a c t a a a a a a a a a a c g a a a c t c g t
30

<210> 158
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 158
c g c c g a g g t c g c t t c c t a a a c
21

<210> 159
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 159
g a g g a a a t a g g c g t g g t t c g
20

<210> 160
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 160

a a a a c c g a a a
25

a a a a c t a a t t

c g t c a

<210> 161

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 161

c c a c g g a c g g
22

c g t t g a c g g t

t c

<210> 162

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 162

c g c g c g g t a t
17

a g t t c g g

<210> 163

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 163

a c c c c c a a c g
17

a a c g a a c

<210> 164
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 164
c g c c g a g g c t a t a c g c g a c g c
21

<210> 165
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 165
g a t t t t g a g g t t a g g t t t c g t c g
23

<210> 166
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 166
g a a a a a c a a c g c t a c t a c c g c
21

<210> 167
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 167

c c a c g g a c g g
22

c g c g g a g t t g

g c

<210> 168

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 168

g t t c g t a t a t
29

a g t t t t a g a a

g t t t a g c g a

<210> 169

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 169

c g t c g a a a a a
25

c t a c c t a a a a

a c g a t

<210> 170

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 170

c g c c g a g g t a
21

c g t a a c t t a a

c

<210> 171
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 171
c g a t t t t g t a g c g t t t g g a t a t c g
24

<210> 172
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 172
t c g a a a a c c c c a c g a a c c
18

<210> 173
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 173
c c a c g g a c g c g t t a a t a c c c c c
22

<210> 174
<211> 28
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 174

t t a a a g g t t t
28

a g t t a a a g a t

g g a a t c g t

<210> 175

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 175

c t c g c g a c g a
19

a a a a a a c c c

<210> 176

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 176

c g c c g a g g c g
21

a a c t c c c a a c

c

<210> 177

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 177

g t g g t t t t a t
24

t t c g t t t c g t

t t c g

<210> 178
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 178
a a a c t a a c g a a t a t a a c t c c g t c g a
25

<210> 179
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 179
c g c c g a g g g c g a c g t t c g a g c
21

<210> 180
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 180
g c g a g a g g a g t a g g t t c g
18

<210> 181
<211> 17
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 181

a a c c t a c g a a
17

c g c g c g a

<210> 182

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 182

c c a c g g a c g a
22

c c c g a c g a c c

a c

<210> 183

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 183

g c g t a c g g t c
17

g c g t a t t

<210> 184

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 184

c a a t a a c t c g
20

a c g a a a c g c g

<210> 185
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 185
c g c c g a g g g c g a a a t t c t a a c
21

<210> 186
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 186
g t t t a g c g c g g g t t t t t c g
19

<210> 187
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 187
c c c g a a c t t c c c g a a c t
17

<210> 188
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 188

c c a c g g a c g g
22

c g g t a g t g g t

c c

<210> 189

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 189

g g a g a a a g g c
17

g g g c g g a

<210> 190

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 190

c g t t c c c a a a
17

a a c g c g a

<210> 191

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 191

c g c c g a g g a c
21

g a c g a c g t a t

c

<210> 192
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 192
c g t t c g t t c g
20

t t c g t t t t g g

<210> 193
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 193
c c c c g c c c c c
17

t c t a a a c

<210> 194
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 194
c c a c g g a c g c
22

g a a c g a a a c g

a c

<210> 195
<211> 18
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 195

g t a g t g g c g g
18

t t a g g t c g

<210> 196

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 196

t c a c c g t c a a
27

t a t t a a t a a a

a a c a c g a

<210> 197

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 197

c g c c g a g g g c
21

g t t g t t t t t t

c

<210> 198

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 198

g c g a g g t t g c
17

g g t t t t g

<210> 199
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 199
c g a a a a a t a a t t c g a t t a a c c a a a a c g c
29

<210> 200
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 200
c c a c g g a c g c c g a t a t t t c c c c
22

<210> 201
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 201
g t t c g t t t t a t a t t t t t c g c g c g
23

<210> 202
<211> 17
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 202

c c g a a c g t c c
17

g c g a a a c

<210> 203

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 203

c g c c g a g g c g
21

a a c t a c t c a a

c

<210> 204

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 204

g g g c g c g a g t
17

a t a g t c g

<210> 205

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 205

c a a c g c g a c t
17

a a t c c g c

<210> 206
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 206
c g c c g a g g c c g t c a c c t c c a c
21

<210> 207
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 207
g g g t c g t a g g a g t g t t t t c g
20

<210> 208
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 208
c c t c t c t a a a a a c c g c t c a a c
21

<210> 209
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 209

c c a c g g a c g c
22

t c g t a a a c g c

c c

<210> 210

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 210

a a g a c g t g g a
19

a a a g t t g c g

<210> 211

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 211

c c g c g c g t c c
17

a t a a t t c

<210> 212

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 212

c c a c g g a c g c
22

g a a a a c t c g t

t c

<210> 213
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 213
g t a g g a g t g g t t g g c g c
17

<210> 214
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 214
c c c a c a c c t c g a c g a a a t
18

<210> 215
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 215
c c a c g g a c g c g c g g a a g t t g a c
22

<210> 216
<211> 17
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 216

g g a g g a g a a c
17

g c g g a t g

<210> 217

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 217

g c g a t c c g c g
18

a a a a a a c g

<210> 218

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 218

c c a c g g a c g g
22

a a c g c c c g a a

a c

<210> 219

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 219

c t a a c t a a a c
29

t a a c g a a t a t

a a c t c c g t c

<210> 220
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 220
g t t a g g g c g g a g a g c g t
17

<210> 221
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 221
c g c g c g a t a a c a a a a a c g
18

<210> 222
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 222
c g c c g a g g g c g a a c t a a a a c c
21

<210> 223
<211> 18
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 223

a c g t t t g g a g
18

a t t c g c g g

<210> 224

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 224

c c c c c g c a a c
16

t t c c t t

<210> 225

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 225

c g c c g a g g g a
21

c g g c g g t t g t

c

<210> 226

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 226

g g g t c g t a g g
18

t g t t c g a a

<210> 227
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 227
g c g c c c t a c t c c t c g a c
17

<210> 228
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 228
c g c c g a g g c g c c g c t a a a c t c
21

<210> 229
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 229
g g t a t c g t a g t t t t t c g c g g a
21

<210> 230
<211> 17
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 230

c t c c t a c g c c
17

g c t c a a a

<210> 231

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 231

c g c c g a g g a t
21

t a g a g g c g a t

c

<210> 232

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 232

g t g t a a a t g c
16

g c g g g c

<210> 233

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 233

c g c t t t a a t t
20

c a a c a c c g c g

<210> 234
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 234
c g c c g a g g c g g t t t a t a c g t c
21

<210> 235
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 235
a g c c g g t t t t c c g g c t g a g a c c t c g g c g c
29

<210> 236
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 236
a g c c g g t t t t c c g g c t g a g a c g t c c g t g g c
30

<210> 237
<211> 30
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 237

a g c c g g t t t t

30

c c g g c t g a g a

c t c c g c g t c c