

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

**(11) Application No. AU 2016315924 B2**

- (54) Title  
**Detecting gastric neoplasm**
- (51) International Patent Classification(s)  
**C12Q 1/68** (2006.01)
- (21) Application No: **2016315924** (22) Date of Filing: **2016.08.31**
- (87) WIPO No: **WO17/040627**
- (30) Priority Data
- (31) Number **62/212,221** (32) Date **2015.08.31** (33) Country **US**
- (43) Publication Date: **2017.03.09**  
(44) Accepted Journal Date: **2022.08.25**
- (71) Applicant(s)  
**Mayo Foundation for Medical Education and Research;Exact Sciences Development Company, LLC**
- (72) Inventor(s)  
**Allawi, Hatim T.;Ahlquist, David A.;Taylor, William R.;Kisiel, John B.;Yab, Tracy C.;Mahoney, Douglas W.**
- (74) Agent / Attorney  
**Spruson & Ferguson, GPO Box 3898, Sydney, NSW, 2001, AU**
- (56) Related Art  
**US 20140274748 A1**  
**WO 2014159652 A2**  
**WO 2016/160454 A1**

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

(19) World Intellectual Property Organization

International Bureau



(10) International Publication Number

WO 2017/040627 A1

(43) International Publication Date

9 March 2017 (09.03.2017)

(51) International Patent Classification:

C12Q 1/68 (2006.01)

(21) International Application Number:

PCT/US2016/049653

(22) International Filing Date:

31 August 2016 (31.08.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/212,221 31 August 2015 (31.08.2015) US

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

(72) Inventors: **ALLAWI, Hatim T.**; 901 Winding Way, Middleton, Wisconsin 53562 (US). **AHLQUIST, David A.**; 200 First Street S.W., Rochester, Minnesota 55905 (US). **TAYLOR, William R.**; 200 First Street S.W., Rochester, Minnesota 55905 (US). **KISIEL, John B.**; 200 First Street S.W., Rochester, Minnesota 55905 (US). **YAB, Tracy C.**; 200 First Street S.W., Rochester, Minnesota 55905 (US). **MAHONEY, Douglas W.**; 200 First Street S.W., Rochester, Minnesota 55905 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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))



WO 2017/040627 A1

(54) Title: DETECTING GASTRIC NEOPLASM

(57) Abstract: Provided herein is technology relating to detecting neoplasia and particularly, but not exclusively, to methods, compositions, and related uses for detecting premalignant and malignant neoplasms such as gastric cancer.

## DETECTING GASTRIC NEOPLASM

### CROSS REFERENCE TO RELATED APPLICATION

This application claims priority to and the benefit of U.S. Provisional Application No. 5 62/212,221, filed August 31, 2016, which is hereby incorporated by reference in its entirety.

### FIELD OF INVENTION

Provided herein is technology relating to detecting neoplasia and particularly, but not exclusively, to methods, compositions, and related uses for detecting premalignant and 10 malignant neoplasms such as gastric cancer.

### BACKGROUND

Gastric cancer is the third most common cause of cancer-related death in the world (see, e.g., World Health Organization. Cancer: Fact Sheet No 297. WHO), and it remains 15 difficult to cure in Western countries, primarily because most patients present with advanced disease. In the United States, stomach malignancy is currently the 15th most common cancer (see, e.g., Surveillance, Epidemiology, and End Results Program. SEER Stat Fact Sheets: Stomach Cancer. National Cancer Institute). In most developed countries, however, rates of stomach cancer have declined dramatically over the past half century.

20 Decreases in gastric cancer have been attributed in part to widespread use of refrigeration, which has had several beneficial effects: increased consumption of fresh fruits and vegetables; decreased intake of salt, which had been used as a food preservative; and decreased contamination of food by carcinogenic compounds arising from the decay of unrefrigerated meat products. Salt and salted foods may damage the gastric mucosa, leading 25 to inflammation and an associated increase in DNA synthesis and cell proliferation. Other factors likely contributing to the decline in stomach cancer rates include lower rates of chronic *Helicobacter pylori* infection, thanks to improved sanitation and use of antibiotics, and increased screening in some countries (see, e.g., Global Cancer Facts & Figures, 3rd ed. American Cancer Society).

30 Nevertheless, gastric cancer remains difficult to cure in Western countries, primarily because most patients present with advanced disease. Even patients who present in the most favorable condition and who undergo curative surgical resection often die of recurrent disease.

Accordingly, improved methods and techniques for early detection of gastric cancer are needed.

## SUMMARY

5 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.

10 DNA methylation may be a more chemically and biologically stable diagnostic tool than RNA or protein expression (Laird (2010) “Principles and challenges of genome-wide DNA methylation analysis” *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) “Highly

15 methylated genes in colorectal neoplasia: implications for screening” *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. 20 (2009) “DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution” *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. 25 (2009) “DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution” *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 (defined as having <5% CpG sequence within a 300 base pair region) tended to be frequently methylated in a dynamic tissue-specific pattern (Meissner 30 et al. (2008) “Genome-scale DNA methylation maps of pluripotent and differentiated cells” *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)

“Genome-scale DNA methylation maps of pluripotent and differentiated cells” *Nature* 454: 766–70).

In aggregate, gastrointestinal cancers account for more cancer mortality than any other organ system. Yet, only colorectal cancers are currently screened domestically. Annual US mortality from upper GI cancers exceeds 90,000 compared to roughly 50,000 for colorectal cancer. Adenocarcinoma of the stomach is the second most frequent cause of cancer death worldwide, with nearly 1 million deaths per year. It is most prevalent in Pacific Asia, Russia, and certain parts of Latin America. Prognosis is poor (5 year survival <5-15%) because most patients present with advanced disease.

The genomic mechanisms underlying stomach cancer are complex and incompletely understood. Chronic gastritis from *Helicobacter pylori* infection along with environmental exposures contribute to the etiology. Chromosomal and microsatellite instability are associated with stomach cancer, as well as epigenetic alterations. A CpG methylator phenotype has also been described in 24%-47% of cases. Commonly observed mutations include the genes p53, APC,  $\beta$ -catenin, and k-ras, but these are relatively infrequent and heterogeneous in nature.

Stomach cancers shed cells and DNA into the digestive stream and are ultimately excreted in stool. Highly sensitive assays have previously been used to detect mutant DNA in matched stools of stomach cancer patients whose excised tumor was known to contain the same sequences. Several limitations with mutation markers, however, relate to the underlying heterogeneity and unwieldy process of their detection; typically, each mutational site across multiple genes must be assayed separately to achieve high sensitivity.

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

Several methods are available to search for novel methylation markers. While 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 promotor. 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 5 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 10 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. This latter approach, termed Reduced Representation Bisulfite Sequencing (RRBS) has, to 15 our knowledge, not yet been used to study stomach cancer.

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 20 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.

Clinical applications of highly discriminant markers could have great impact. For 25 example, assay of such markers in distant media like stool or blood could lead to accurate screening or diagnostic tools for detection of gastric neoplasia.

Experiments conducted during the course of developing embodiments for the present technology list and describe 123 novel DNA methylation markers generated in a similar fashion from stomach cancer samples. Along with the cancers, normal stomach tissues, 30 normal colon tissues, and normal white blood cell DNA were sequenced. The markers were validated on a larger sample set to identify and optimize the most robust candidates. Additional experiments identified 10 optimal markers for detecting stomach cancer (ARHGEF4, ELMO1, ABCB1, CLEC11A, ST8SIA1, SFMBT2, CD1D, CYP26C1, ZNF569, and C13ORF18; see Example 2 and Tables 4, 5 and 6). Additional experiments identified 30

optimal markers for detecting stomach cancer (ELMO1, ARHGEF4, EMX1, SP9, CLEC11A, ST8SIA1, BMP3, KCNA3, DMRTA2, KCNK12, CD1D, PRKCB, CYP26C1, ZNF568, ABCB1, ELOVL2, PKIA, SFMBT2 (893), PCBP3, MATK, GRN2D, NDRG4, DLX4, PPP2R5C, FGF14, ZNF132, CHST2 (7890), FLI1, c13orf18, or ZNF569; see Example 1 and 5 Tables 1, 2 and 3). Additional experiments identified 12 optimal markers for detecting gastric cancer (e.g., stomach cancer) in plasma samples (ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP2R5C, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1; see Example 3 and Tables 7 and 8).

Accordingly, provided herein is technology for stomach cancer screening markers that 10 provide a high signal-to-noise ratio and a low background level when detected from samples taken from a subject (e.g., stool sample; gastric tissue sample; plasma sample). Markers were identified in a case-control study by comparing the methylation state of DNA markers from gastric tissue or plasma of subjects with stomach cancer to the methylation state of the same DNA markers from control subjects (e.g., normal stomach tissue, normal colonic epithelia, 15 normal plasma, and normal white blood cell derived DNA; see, Examples 1, 2 and 3, and Tables 1-8) (e.g., plasma tissue; see, Example 3).

As described herein, the technology provides a number of methylated DNA markers and subsets thereof (e.g., sets of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more markers) with high discrimination for stomach cancer overall. Experiments applied a selection filter to candidate 20 markers to identify markers that provide a high signal to noise ratio and a low background level to provide high specificity for purposes of stomach cancer screening or diagnosis.

In some embodiments, the technology is related to assessing the presence of and 25 methylation state of one or more of the markers identified herein in a biological sample (e.g., gastric tissue, plasma sample). These markers comprise one or more differentially methylated regions (DMR) as discussed herein, e.g., as provided in Tables 1, 2, 4 and/or 7. Methylation state is assessed in embodiments of the technology. As such, the technology provided herein is not restricted in the method by which a gene's methylation state is measured. For example, in some embodiments the methylation state is measured by a genome scanning method. For example, one method involves restriction landmark genomic scanning (Kawai et al. (1994) 30 *Mol. Cell. Biol.* 14: 7421–7427) and another example involves methylation-sensitive arbitrarily primed PCR (Gonzalgo et al. (1997) *Cancer Res.* 57: 594–599). In some embodiments, changes in methylation patterns at specific CpG sites are monitored by digestion of genomic DNA with methylation-sensitive restriction enzymes followed by Southern analysis of the regions of interest (digestion-Southern method). In some

embodiments, analyzing changes in methylation patterns involves a PCR-based process that involves digestion of genomic DNA with methylation-sensitive restriction enzymes 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 5 methylation analysis. These include methylation-specific PCR (MSP) (Herman et al. (1992) *Proc. Natl. Acad. Sci. USA* 93: 9821–9826) and restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA (Sadri and Hornsby (1996) *Nucl. Acids Res.* 24: 5058–5059; and Xiong and Laird (1997) *Nucl. Acids Res.* 25: 2532–2534). PCR techniques have been developed for detection of gene mutations (Kuppuswamy et al. (1991) *Proc. Natl. 10 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 15 “quantitative Ms-SNuPE assay” as described in U.S. Pat. No. 7,037,650 are used in some embodiments.

Upon evaluating a methylation state, the methylation state is often expressed as the fraction or percentage of individual strands of DNA that is methylated at a particular site (e.g., at a single nucleotide, at a particular region or locus, at a longer sequence of interest, e.g., up to a ~100-bp, 200-bp, 500-bp, 1000-bp subsequence of a DNA or longer) relative to 20 the total population of DNA in the sample comprising that particular site. Traditionally, the amount of the unmethylated nucleic acid is determined by PCR using calibrators. Then, a known amount of DNA is bisulfite treated and the resulting methylation-specific sequence is determined using either a real-time PCR or other exponential amplification, e.g., a QuARTS assay (e.g., as provided by U.S. Pat. No. 8,361,720; and U.S. Pat. Appl. Pub. Nos. 25 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 30 quantitative standards. Then, a second standard curve for the methylated target is constructed from at least two points and external standards. This second standard curve relates the Ct for methylated DNA to known quantitative standards. Next, the test sample Ct values are determined for the methylated and unmethylated populations and the genomic equivalents of DNA are calculated from the standard curves produced by the first two steps. The percentage of methylation at the site of interest is calculated from the amount of methylated DNAs

relative to the total amount of DNAs in the population, e.g., (number of methylated DNAs) / (the number of methylated DNAs + number of unmethylated DNAs) × 100.

Also provided herein are compositions and kits for practicing the methods. For example, in some embodiments, reagents (e.g., primers, probes) specific for one or more markers are provided alone or in sets (e.g., sets of primers pairs for amplifying a plurality of markers). Additional reagents for conducting a detection assay may also be provided (e.g., enzymes, buffers, positive and negative controls for conducting QuARTS, PCR, sequencing, bisulfite, or other assays). In some embodiments, the kits 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 1-274 as provided in Tables 1, 2, 4 and 7); comparing methylation states (e.g., of one or more DMR, e.g., DMR 1-274 as provided in Tables 1, 2, 4 and 7); 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 1-274 as provided in Tables 1, 2, 4 and 7); 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, 2, 4 and 7). 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, 2, 4 and 7). In some embodiments, the methylation state of a DMR defines a dimension and may have values in a 5 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 10 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 15 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 20 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 25 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 30 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.

10 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, 15 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, 20 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, 25 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 30 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 a neoplasm 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., gastric tissue) (e.g., plasma sample) and identifying the subject as having a neoplasm when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have a neoplasm, wherein the marker comprises a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1–274 as provided in Tables 1, 2, 4 and/or 5 7. The technology is related to identifying and discriminating gastric cancer, e.g., in some embodiments the neoplasm is a gastric neoplasm (e.g., stomach cancer). Some embodiments provide methods comprising assaying a plurality of markers, e.g., comprising assaying 10 2 to 11 to 100 or 200 or 274 markers.

The technology is not limited in the methylation state assessed. In some embodiments 15 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 20 methylation of the marker relative to a normal methylation state of the marker. In some embodiments, the methylation state of the marker comprises a decreased methylation of the marker relative to a normal methylation state of the marker. In some embodiments the methylation state of the marker comprises a different pattern of methylation of the marker relative to a normal methylation state of the marker.

Furthermore, in some embodiments the marker is a region of 100 or fewer bases, the 25 marker is a region of 500 or fewer bases, the marker is a region of 1000 or fewer bases, the marker is a region of 5000 or fewer bases, or, in some embodiments, the marker is one base. In some embodiments the marker is in a high CpG density promoter.

The technology is not limited by sample type. For example, in some embodiments the sample is a stool sample, a tissue sample (e.g., gastric tissue sample), a blood sample (e.g., 30 plasma, serum, whole blood), an excretion, or a urine sample.

Furthermore, the technology is not limited in the method used to determine methylation state. In some embodiments the assaying comprises using methylation specific polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation specific nuclease, mass-based separation, or target capture. In some embodiments, the assaying

comprises use of a methylation specific oligonucleotide. In some embodiments, the technology uses massively parallel sequencing (e.g., next-generation sequencing) to determine methylation state, e.g., sequencing-by-synthesis, real-time (e.g., single-molecule) sequencing, bead emulsion sequencing, nanopore sequencing, etc.

5 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: 1–109. In some embodiments are provided an oligonucleotide comprising a sequence complementary to a chromosomal region having a base in a DMR, e.g., an oligonucleotide sensitive to methylation state of a DMR.

10 The technology provides various panels of markers, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ARHGEF4, ELMO1, ABCB1, CLEC11A, ST8SIA1, SFMBT2, CD1D, CYP26C1, ZNF569, or C13ORF18, and that comprises the marker (see, Table 4). In addition, embodiments provide a method of analyzing a DMR from Table 4 that is DMR No. 49, 43, 196, 66, 1, 237, 249, 250, 251 or 252. Some embodiments provide determining the methylation state of a marker, wherein a chromosomal region having an annotation that is ELMO1, ARHGEF4, EMX1, SP9, CLEC11A, ST8SIA1, BMP3, KCNA3, DMRTA2, KCNK12, CD1D, PRKCB, CYP26C1, ZNF568, ABCB1, ELOVL2, PKIA, SFMBT2 (893), PCBP3, MATK, GRN2D, NDRG4, DLX4, PPP2R5C, FGF14, ZNF132, CHST2 (7890), FLI1, c13orf18, or ZNF569, and that comprises the marker (see, Table 2). In addition, embodiments provide a method of analyzing a DMR from Table 2 that is DMR No. 253, 251, 254, 255, 256, 249, 257, 258, 259, 260, 261, 262, 250, 263, 1, 264, 265, 196, 266, 118, 267, 268, 269, 270, 271, 272, 46, 273, 252, or 237. In some embodiments wherein the obtained sample is a plasma sample, the marker comprises a chromosomal region having an annotation that is ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1, and that comprises the marker. In addition, in such embodiments wherein the sample is a plasma sample, a method of analyzing a DMR from Tables 1, 2, 4 and 7 that is DMR No. 253, 251, 254, 255, 256, 249, 257, 258, 259, 260, 261, 262, 250, 263, 1, 264, 265, 196, 266, 118, 267, 268, 269, 270, 271, 272, 46, 273, 252, or 237. In some embodiments, the methods comprise 20 determining the methylation state of two markers, e.g., a pair of markers provided in a row of Tables 1, 2, 4 or 7.

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 1–274 (from Tables 1, 2, 4 or 7) and having a methylation state associated with a subject who

does not have a cancer (e.g., stomach cancer). In some embodiments, kits comprise a bisulfite reagent and an 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 1–274 (from Tables 1, 2, 4 or 7) and having a methylation state associated with a subject who has a cancer (e.g., stomach cancer). Some kit embodiments comprise a sample collector for obtaining a sample from a subject (e.g., a stool sample; gastric tissue sample; plasma 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).

10 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 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 a neoplasm in a sample obtained from a subject (e.g., gastric 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 1–274 (from Tables 1, 2, 4 or 7); 20 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 a 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, 25 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 30 DMR from a subject who does not have a cancer to identify differences in the two sequences; and identifying the subject as having a neoplasm when a difference is present.

Systems for screening for a neoplasm in a sample obtained from a subject are provided by the technology. Exemplary embodiments of systems include, e.g., a system for screening for a neoplasm in a sample obtained from a subject (e.g., gastric 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 cancer-

5        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, 2, or 7) 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  
10      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.

15        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  
20      who do not have a 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 a cancer. Related system embodiments comprise a set of nucleic acids as described and a database of nucleic acid sequences associated with the set of nucleic acids. Some embodiments further  
25      comprise a bisulfite reagent. And, some embodiments further comprise a nucleic acid sequencer.

30        In certain embodiments, methods for characterizing a sample (e.g., gastric tissue sample; plasma sample; stool sample) from a human patient are provided. For example, in some embodiments such embodiments comprise obtaining DNA from a sample of a human patient; assaying a methylation state of a DNA methylation marker comprising a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1-274 from Tables 1, 2, or 7; and comparing the assayed methylation state of the one or more DNA methylation markers with methylation level references for the one or more DNA methylation markers for human patients not having gastric neoplasia.

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

5 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 107 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 10 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.

15 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 20 methylation marker is in a high CpG density promoter.

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

25 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: 1-109.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ARHGEF4, ELMO1, ABCB1, CLEC11A, ST8SIA1, SFMBT2, CD1D, CYP26C1, ZNF569, and C13ORF18 comprises the DNA methylation marker. In some 30 embodiments, the DMR is from Table 4 and is selected from a group consisting of DMR No. 49, 43, 196, 66, 1, 237, 249, 250, 251 and 252.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ELMO1, ARHGEF4, EMX1, SP9, CLEC11A, ST8SIA1, BMP3, KCNA3, DMRTA2, KCNK12, CD1D, PRKCB, CYP26C1, ZNF568, ABCB1, ELOVL2,

PKIA, SFMBT2 (893), PCBP3, MATK, GRN2D, NDRG4, DLX4, PPP2R5C, FGF14, ZNF132, CHST2 (7890), FLI1, c13orf18, and ZNF569 comprises the DNA methylation marker.

In some embodiments, the DMR is from Table 2 and is selected from a group consisting of DMR No. 253, 251, 254, 255, 256, 249, 257, 258, 259, 260, 261, 262, 250, 263, 1, 264, 265, 196, 266, 118, 267, 268, 269, 270, 271, 272, 46, 273, 252, and 237.

In some embodiments wherein the obtained sample is a plasma sample, the marker comprises a chromosomal region having an annotation that is ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and 10 ST8SIA1, and that comprises the marker.

In some embodiments, the DMR is from Table 1, 2, 4 or 7 and is selected from a group consisting of DMR No. 253, 237, 252, 261, 251, 196, 250, 265, 256, 249, and 274.

In some embodiments, such methods comprise determining the methylation state of two DNA methylation markers. In some embodiments, such methods comprise determining 15 the methylation state of a pair of DNA methylation markers provided in a row of Tables 1, 2, 4 or 7.

In certain embodiments, the technology provides methods for characterizing a sample obtained from a human patient. In some embodiments, such methods comprise determining a methylation state of a DNA methylation marker in the sample comprising a base in a DMR 20 selected from a group consisting of DMR 1–274 from Tables 1, 2, 4 and 7; 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 gastric 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 25 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., gastric tissue sample; plasma sample; stool sample), the method comprising reacting a nucleic acid comprising a DMR with a bisulfite reagent to 30 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 a gastric 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., gastric 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 5 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 gastric 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 10 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 gastric tissue sample, a blood sample, or a urine sample.

In some embodiments, the database comprises nucleic acid sequences comprising a 15 DMR. In some embodiments, the database comprises nucleic acid sequences from subjects who do not have a gastric cancer.

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

## 20 BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present technology will become better understood with regard to the following drawings:

FIG. 1 shows that ELMO1 exhibited high discrimination for gastric cancer as described in Example 2.

25 Fig. 2 provides the oligonucleotide sequences for the FRET cassettes, used in the detection of methylated DNA signatures by QuARTs (quantitative allele-specific real-time target and signal amplification) assay. Each FRET sequence includes a fluorophore and quencher which can be multiplexed together into 3 separate assays.

30 Fig. 3 provides receiver operating characteristic curve highlighting the performance in plasma of the 3 marker panel (ELMO1, ZNF569, and c13orf18) compared with individual marker curves. At 100% specificity, the panel detected gastric cancer at a sensitivity of 86%.

Fig. 4A provides log scale absolute strand count of methylated ELMO1 in plasma by stage illustrating the progression from normal mucosa to stage 4 gastric cancer. Quantitative marker levels increased directly with GC stage.

Fig. 4B provides a bar chart demonstrating gastric cancer sensitivity in plasma by stage (at 100% specificity) of the 3 marker panel (ELMO1, ZNF569, and c13orf18).

Fig. 5A, B and C provides performance of 12 gastric cancer markers (ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, 5 LRRC4, and ST8SIA1; ELMO1 was run twice, the second time in a bplex format). in 74 plasma samples in a matrix format at 90% specificity (A), 95% specificity (B), and 100% specificity (C). Markers are listed vertically and samples horizontally. Samples are arranged with normals on the left and cancers on the right. Positive hits are in light grey and misses in dark grey. Here, the top performing marker – ELMO1 – is listed first at 90% specificity and 10 the remaining panel at 100% specificity. This plot allows markers to be assessed in a complementary fashion.

Fig. 6 provides performance of the 3 gastric cancer markers (final panel) in 74 plasma samples at 100% specificity. Markers are listed vertically and samples horizontally. Samples are arranged with normals on the left and cancers on the right. The cancers are arranged by 15 stage. Positive hits are in light grey and misses in dark grey (ELMO1, ZNF569, and c13orf18).

Fig. 7A-M provides box plots (linear scale) of 12 gastric cancer markers in plasma (ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1; ELMO1 was run twice, the second time in a bplex format). Control samples (N=38) are listed on the left and gastric cancer cases (N=36) on the 20 right. The vertical axis is % methylation normalized to  $\beta$ -actin strands.

It is to be understood that the figures are not necessarily drawn to scale, nor are the objects in the figures necessarily drawn to scale in relationship to one another. The figures are 25 depictions that are intended to bring clarity and understanding to various embodiments of apparatuses, systems, compositions, and methods disclosed herein. Wherever possible, the same reference numbers are used throughout the drawings to refer to the same or like parts. Moreover, it should be appreciated that the drawings are not intended to limit the scope of the present teachings in any way.

### 30 DETAILED DESCRIPTION

Provided herein is technology relating to detecting neoplasia and particularly, but not exclusively, to methods, compositions, and related uses for detecting premalignant and malignant neoplasms such as gastric cancer. As the technology is described herein, the

section headings used are for organizational purposes only and are not to be construed as limiting the subject matter in any way.

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

5 embodiments disclosed. One skilled in the art will appreciate, however, that these various embodiments may be practiced with or without these specific details. In other instances, structures and devices are shown in block diagram form. Furthermore, one skilled in the art can readily appreciate that the specific sequences in which methods are presented and performed are illustrative and it is contemplated that the sequences can be varied and still 10 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.

## 20 **Definitions**

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

Throughout the specification and claims, the following terms take the meanings 25 explicitly associated herein, unless the context clearly dictates otherwise. The phrase “in one embodiment” as used herein does not necessarily refer to the same embodiment, though it may. Furthermore, the phrase “in another embodiment” as used herein does not necessarily refer to a different embodiment, although it may. Thus, as described below, various embodiments of the invention may be readily combined, without departing from the scope or 30 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 5 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 10 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 15 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 20 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 25 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 30 characteristics of a gene isolated from a naturally occurring source. The term “wild-type” when made in reference to a gene product refers to a gene product that has the characteristics of a gene product isolated from a naturally occurring source. The term “naturally-occurring” as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that

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

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

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

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

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. al., *Nucleic Acids Research*, 25:1854-1858 (1997); U.S. Patent No. 5,508,169; each of which are herein incorporated by reference in their entireties), methylation-specific PCR (see, e.g., Herman, et al., (1996) *PNAS* 93(13) 9821-9826; herein incorporated by reference in its entirety), miniprimer PCR, multiplex ligation-dependent probe amplification (see, e.g., Schouten, et al., (2002) *Nucleic Acids Research* 30(12): e57; herein incorporated by reference in its entirety), multiplex PCR (see, e.g., Chamberlain, et al., (1988) *Nucleic Acids Research* 16(23) 11141-11156; Ballabio, et al., (1990) *Human Genetics* 84(6) 571-573; Hayden, et al., (2008) *BMC Genetics* 9:80; each of which are herein incorporated by reference in their entireties), nested PCR, overlap-extension PCR (see, e.g., Higuchi, et al., (1988) *Nucleic Acids Research* 16(15) 7351-7367; herein incorporated by reference in its entirety), real time PCR (see, e.g., Higuchi, et al. al., (1992) *Biotechnology* 10:413-417; Higuchi, et al., (1993) *Biotechnology* 11:1026-1030; each of which are herein incorporated by reference in their entireties), reverse transcription PCR (see, e.g., Bustin, S.A. (2000) *J. Molecular Endocrinology* 25:169-193; herein incorporated by reference in its entirety), solid phase PCR, thermal asymmetric interlaced PCR, and Touchdown PCR (see, e.g., Don, et al., *Nucleic Acids Research* (1991) 19(14) 4008; Roux, K. (1994) *Biotechniques* 16(5) 812-814; Hecker, et al., (1996) *Biotechniques* 20(3) 478-485; each of which are herein incorporated by reference in their entireties). Polynucleotide amplification also can be accomplished using digital PCR (see, e.g., Kalinina, et al., *Nucleic Acids Research* 25; 1999-2004, (1997); Vogelstein and Kinzler, *Proc Natl Acad Sci USA*. 96; 9236-41, (1999); International Patent Publication No. WO05023091A2; US Patent Application Publication No. 20070202525; each of which are incorporated herein by reference in their entireties).

The term “polymerase chain reaction” (“PCR”) refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of

introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the

5 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

10 one “cycle”; there can be numerous “cycles”) to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each

other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the “polymerase chain reaction” (“PCR”). Because

15 the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be “PCR amplified” and are “PCR

products” or “amplicons.”

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.

20 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

25 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

30 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.”

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

The term “primer” refers to an oligonucleotide, whether occurring naturally as 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 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

5 separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer, and the use of the method.

10 The term “probe” refers to an oligonucleotide (e.g., a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly, or by PCR amplification, that is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification, and isolation of particular gene sequences (e.g., a

15 “capture probe”). It is contemplated that any probe used in the present invention may, in some embodiments, be labeled with any “reporter molecule,” so that is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

20 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

25 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 5 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 10 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 15 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 20 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 25 of methylated and unmethylated nucleotides are different. Sequences are said to be “differentially methylated” or as having a “difference in methylation” or having a “different methylation state” when they differ in the extent (e.g., one has increased or decreased methylation relative to the other), frequency, or pattern of methylation. The term “differential methylation” refers to a difference in the level or pattern of nucleic acid methylation in a 30 cancer positive sample as compared with the level or pattern of nucleic acid methylation in a cancer negative sample. It may also refer to the difference in levels or patterns between patients that have recurrence of cancer after surgery versus patients who not have recurrence. Differential methylation and specific levels or patterns of DNA methylation are prognostic

and predictive biomarkers, e.g., once the correct cut-off or predictive characteristics have been defined.

Methylation state frequency can be used to describe a population of individuals or a sample from a single individual. For example, a nucleotide locus having a methylation state frequency of 50% is methylated in 50% of instances and unmethylated in 50% of instances.

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

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

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

As used herein, a “CpG island” refers to a G:C-rich region of genomic DNA containing an increased number of CpG dinucleotides relative to total genomic DNA. A CpG island can be at least 100, 200, or more base pairs in length, where the G:C content of the region is at least 50% and the ratio of observed CpG frequency over expected frequency is 0.6; in some instances, a CpG island can be at least 500 base pairs in length, where the G:C content of the region is at least 55%) and the ratio of observed CpG frequency over expected frequency is 0.65. The observed CpG frequency over expected frequency can be calculated according to the method provided in Gardiner-Garden et al (1987) *J. Mol. Biol.* **196**: 261–281. For example, the observed CpG frequency over expected frequency can be calculated according to the formula  $R = (A \times B) / (C \times D)$ , 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 5 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 10 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, 15 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 20 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 25 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 30 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.

5 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.

10 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.

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

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

20 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.

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

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

30 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 nucleotide refers specifically to an unmethylated typically occurring nucleotide.

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 5 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 10 instance a recognition sequence such as CGCG or CCCGGG). Further preferred for some embodiments are restriction enzymes that do not cut if the cytosine in this dinucleotide is methylated at the carbon atom C5.

As used herein, a “different nucleotide” refers to a nucleotide that is chemically different from a selected nucleotide, typically such that the different nucleotide has Watson- 15 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 20 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 25 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 30 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 5 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 10 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 15 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 20 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 25 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, 30 “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 5 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 10 the presence or likelihood of a subject contracting a neoplasm or the likelihood that such a subject will respond favorably to a compound (e.g., a pharmaceutical, e.g., a drug) or other treatment.

The term “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 15 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 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 20 found in nature. In contrast, non-isolated nucleic acids, such as DNA and RNA, are found in the state they exist in nature. Examples of non-isolated nucleic acids include: a given DNA sequence (e.g., a gene) found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. 25 However, isolated nucleic acid encoding a particular protein includes, by way of example, such nucleic acid in cells ordinarily expressing the protein, where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or oligonucleotide may be present in single-stranded or double-stranded form. When an isolated 30 nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide may be single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide may be double-stranded). An isolated nucleic acid may, after isolation from its natural or typical environment, be combined with other nucleic acids or molecules. For example, an isolated

nucleic acid may be present in a host cell in which into which it has been placed, e.g., for heterologous expression.

The term “purified” refers to molecules, either nucleic acid or amino acid sequences that are removed from their natural environment, isolated, or separated. An “isolated nucleic acid sequence” may therefore be a purified nucleic acid sequence. “Substantially purified” molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated. As used herein, the terms “purified” or “to purify” also refer to the removal of contaminants from a sample.

The removal of contaminating proteins results in an increase in the percent of polypeptide or nucleic acid of interest in the sample. In another example, recombinant polypeptides are expressed in plant, bacterial, yeast, or mammalian host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

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

The term “sample” is used in its broadest sense. In one sense it can refer to an animal cell or tissue. In another sense, it is meant to include a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. In some embodiments, the sample is a plasma sample. In some embodiments, the sample is a gastric 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 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 pancreas), the sample is a remote sample.

As used herein, the terms “patient” or “subject” refer to organisms to be subject to various tests provided by the technology. The term “subject” includes animals, preferably

mammals, including humans. In a preferred embodiment, the subject is a primate. In an even more preferred embodiment, the subject is a human.

As used herein, the term “kit” refers to any delivery system for delivering materials. In the context of reaction assays, such delivery systems include systems that allow for the 5 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 10 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 15 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 20 fragmented and combined kits.

### **Embodiments of the technology**

Provided herein is technology relating to detecting neoplasia and particularly, but not exclusively, to methods, compositions, and related uses for detecting premalignant and 25 malignant neoplasms such as gastric cancer. Markers were identified in a case-control study by comparing the methylation state of DNA markers from tumors of subjects with gastric cancer (e.g., stomach cancer) to the methylation state of the same DNA markers from control subjects (see, Examples 1 and 2). Additional experiments identified 10 optimal markers for detecting stomach cancer (ARHGEF4, ELMO1, ABCB1, CLEC11A, ST8SIA1, SFMBT2, 30 CD1D, CYP26C1, ZNF569, and C13ORF18; see Example 2 and Table 4). Additional experiments identified 30 optimal markers for detecting stomach cancer (see Example 1 and Table 2). Additional experiments identified 12 optimal markers for detecting gastric cancer (e.g., stomach cancer) in plasma samples (ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4,

SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1; see Example 3 and Tables 1, 2, 4 and 7).

Markers and/or panels of markers (e.g., a chromosomal region having an annotation selected from ARHGEF4, ELMO1, ABCB1, CLEC11A, ST8SIA1, SFMBT2, CD1D, CYP26C1, ZNF569, or C13ORF18, and that comprises the marker (see, Table 4)) were identified in a case-control study by comparing the methylation state of DNA markers (e.g., from gastric tissue of subjects with gastric cancer (e.g., stomach cancer) to the methylation state of the same DNA markers from control subjects (see, Example 2).

Markers and/or panels of markers (e.g., a chromosomal region having an annotation selected from ELMO1, ARHGEF4, EMX1, SP9, CLEC11A, ST8SIA1, BMP3, KCNA3, DMRTA2, KCNK12, CD1D, PRKCB, CYP26C1, ZNF568, ABCB1, ELOVL2, PKIA, SFMBT2 (893), PCBP3, MATK, GRN2D, NDRG4, DLX4, PPP2R5C, FGF14, ZNF132, CHST2 (7890), FLI1, c13orf18, or ZNF569, and that comprises the marker (see, Table 2)) were identified in a case-control study by comparing the methylation state of DNA markers (e.g., from gastric tissue of subjects with gastric cancer (e.g., stomach cancer) to the methylation state of the same DNA markers from control subjects (see, Example 1).

Markers and/or panels of markers (e.g., a chromosomal region having an annotation selected from ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1, and that comprises the marker (see, Tables 1, 2, 4 and 7)) were identified in a case-control study by comparing the methylation state of DNA markers (e.g., from plasma of subjects with gastric cancer (e.g., stomach cancer) to the methylation state of the same DNA markers from control subjects (see, Example 3).

In addition, the technology provides various panels of markers, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ARHGEF4, ELMO1, ABCB1, CLEC11A, ST8SIA1, SFMBT2, CD1D, CYP26C1, ZNF569, or C13ORF18, and that comprises the marker (see, Table 4). In addition, embodiments provide a method of analyzing a DMR from Table 4 that is DMR No. 49, 43, 196, 66, 1, 237, 249, 250, 251 or 252. Some embodiments provide determining the methylation state of a marker, wherein a chromosomal region having an annotation that is ELMO1, ARHGEF4, EMX1, SP9, CLEC11A, ST8SIA1, BMP3, KCNA3, DMRTA2, KCNK12, CD1D, PRKCB, CYP26C1, ZNF568, ABCB1, ELOVL2, PKIA, SFMBT2 (893), PCBP3, MATK, GRN2D, NDRG4, DLX4, PPP2R5C, FGF14, ZNF132, CHST2 (7890), FLI1, c13orf18, or ZNF569, and that comprises the marker (see, Table 2). In addition, embodiments provide a method of

analyzing a DMR from Table 2 that is DMR No. 253, 251, 254, 255, 256, 249, 257, 258, 259, 260, 261, 262, 250, 263, 1, 264, 265, 196, 266, 118, 267, 268, 269, 270, 271, 272, 46, 273, 252, or 237. In some embodiments wherein the obtained sample is a plasma sample, the marker comprises a chromosomal region having an annotation that is ELMO1, ZNF569, 5 C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1, and that comprises the marker. In addition, in such embodiments wherein the sample is a plasma sample, a method of analyzing a DMR from Tables 1, 2, 4 and 7 that is DMR No. 253, 251, 254, 255, 256, 249, 257, 258, 259, 260, 261, 262, 250, 263, 1, 264, 265, 196, 266, 118, 267, 268, 269, 270, 271, 272, 46, 273, 252, or 237. In some embodiments, the 10 methods comprise determining the methylation state of two markers, e.g., a pair of markers provided in a row of Tables 1, 2, 4 or 7.

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.

15 In particular aspects, the present technology provides compositions and methods for identifying, determining, and/or classifying a cancer such as gastric cancer (e.g., stomach 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, gastric tissue sample, plasma sample), wherein a change in the methylation state of the marker is indicative 20 of the presence, class, or site of a gastric cancer (e.g., stomach cancer). Particular embodiments relate to markers comprising a differentially methylated region (DMR, e.g., DMR 1–274, see Tables 1, 2, 4 and 7) that are used for diagnosis (e.g., screening) of neoplastic cellular proliferative disorders (e.g., cancer), including early detection during the pre-cancerous stages of disease.

25 Furthermore, the markers are used for the differentiation of neoplastic from benign cellular proliferative disorders. In particular aspects, the present technology discloses a method wherein a neoplastic cell proliferative disorder is distinguished from a benign cell proliferative disorder.

30 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 1–274) provided herein and listed in Tables 1, 2, 4 or 7 is analyzed, the technology also provides panels of markers comprising at least one marker, region of a marker, or base of a marker comprising a DMR with utility for the detection of cancers, in particular gastric 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.

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 1-274, see Tables 1, 2, 4 and 7). Genomic CpG dinucleotides can be methylated or unmethylated (alternatively known as up- and down-methylated respectively). However the methods of the present invention are suitable for the analysis of biological samples of a heterogeneous nature, e.g., a low concentration of tumor cells, or biological materials therefrom, within a background of a remote sample (e.g., blood, organ effluent, or stool). Accordingly, when analyzing the methylation status of a CpG position within such a sample one may use a quantitative assay for determining the level (e.g., percent, fraction, ratio, proportion, or degree) of methylation at a particular CpG position.

According to the present technology, determination of the methylation status of CpG dinucleotide sequences in markers comprising a DMR has utility both in the diagnosis and characterization of cancers such as gastric 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 (e.g., DMR Nos. 1-248) or Table 4 (e.g., DMR Nos., 49, 43, 196, 66, 1, 237, 249, 250, 251 or 252) or Table 2 (DMR Nos. 253, 251, 254, 255, 256, 249, 257, 258, 259, 260, 261, 262, 250, 263, 1, 264, 265, 196, 266, 118, 267, 268, 269, 270, 271, 272, 46, 273, 252, or 237) or Table 7 (DMR No. 274), or more markers comprising a DMR. 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., gastric cancer). In some embodiments, a marker or a combination of markers discriminates between types and/or locations of a neoplasm.

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.

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

387–95; Feil et al. (1994) *Nucleic Acids Res.* 22: 695; Martin et al. (1995) *Gene* 157: 261–4; WO 9746705; WO 9515373).

Various methylation assay procedures are known in the art and can be used in conjunction with bisulfite treatment according to the present technology. These assays allow 5 for determination of the methylation state of one or a plurality of CpG dinucleotides (e.g., CpG islands) within a nucleic acid sequence. Such assays involve, among other techniques, sequencing of bisulfite-treated nucleic acid, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction enzymes.

For example, genomic sequencing has been simplified for analysis of methylation 10 patterns and 5-methylcytosine distributions by using bisulfite treatment (Frommer et al. (1992) *Proc. Natl. Acad. Sci. USA* 89: 1827–1831). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA finds use in assessing methylation state, e.g., as described by Sadri & Hornsby (1997) *Nucl. Acids Res.* 24: 5058–5059 or as embodied in the method known as COBRA (Combined Bisulfite Restriction 15 Analysis) (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532–2534).

COBRA™ analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific loci in small amounts of genomic DNA (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the 20 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 25 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.

30 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 (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997), methylation-specific PCR (“MSP”; Herman et al., Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; U.S. Pat. No. 5,786,146), and methylated CpG island amplification (“MCA”; Toyota et al., Cancer Res. 59:2307-12, 1999) are used alone or in combination with one or more of these methods.

The “HeavyMethyl™” assay, technique is a quantitative method for assessing methylation differences based on methylation-specific amplification of bisulfite-treated DNA. Methylation-specific blocking probes (“blockers”) covering CpG positions between, or covered by, the amplification primers enable methylation-specific selective amplification of a nucleic acid sample.

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

Typical reagents (e.g., as might be found in a typical MethyLight™-based kit) for HeavyMethyl™ analysis may include, but are not limited to: PCR primers for specific loci (e.g., specific genes, markers, 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.

5 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 10 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.

15 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 20 provided by a reaction in which neither the primers, nor the probe, overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing the biased PCR pool with either control oligonucleotides that do not cover known methylation sites (e.g., a fluorescence-based version of the HeavyMethyl™ and MSP techniques) or with oligonucleotides covering potential methylation sites.

25 The MethyLight™ process is used with any suitable probe (e.g. a “TaqMan®” probe, a Lightcycler® probe, etc.) For example, in some applications double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes, e.g., with MSP primers and/or HeavyMethyl blocker oligonucleotides and a TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent “reporter” and 30 “quencher” molecules and is designed to be specific for a relatively high GC content region so that it melts at about a 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5' to

3' endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Typical reagents (e.g., as might be found in a typical MethylLight™-based kit) for 5 MethylLight™ 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 QM™ (quantitative methylation) assay is an alternative quantitative test for 10 methylation patterns in genomic DNA samples, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe, overlie any CpG dinucleotides.

15 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, 20 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 25 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 30 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 (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site of interest. Small amounts of DNA can be analyzed (e.g., microdissected pathology sections) and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

Typical reagents (e.g., as might be found in a typical Ms-SNuPE™-based kit) for Ms-SNuPE™ analysis may include, but are not limited to: PCR primers for specific loci (e.g., specific genes, markers, 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.

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

A typical protocol for RRBS comprises the steps of digesting a nucleic acid sample with a restriction enzyme such as MspI, filling in overhangs and A-tailing, ligating adaptors,

bisulfite conversion, and PCR. See, e.g., et al. (2005) “Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution” *Nat Methods* 7: 133–6; Meissner et al. (2005) “Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis” *Nucleic Acids Res.* 33: 5868–77.

5 In some embodiments, a quantitative allele-specific real-time target and signal amplification (QuARTS) assay is used to evaluate methylation state. Three reactions sequentially occur in each QuARTS assay, including amplification (reaction 1) and target probe cleavage (reaction 2) in the primary reaction; and FRET cleavage and fluorescent signal generation (reaction 3) in the secondary reaction. When target nucleic acid is amplified  
10 with specific primers, a specific detection probe with a flap sequence loosely binds to the amplicon. The presence of the specific invasive oligonucleotide at the target binding site causes cleavase 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  
15 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  
20 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  
25 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  
30 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 1–274 as provided in Tables 1, 2, 4 and 7) 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

5 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 gastric 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 1-274, e.g., as provided in Tables 1, 2, 4 and 7) and
- 2) detecting a neoplasm or proliferative disorder (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

15

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 gastric tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ARHGEF4, ELMO1, ABCB1, CLEC11A, ST8SIA1, SFMBT2, CD1D, CYP26C1, ZNF569, and C13ORF18, and
- 2) detecting gastric 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:

30

- 1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from a body fluids such as a stool sample or gastric tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having

an annotation selected from the group consisting of ELMO1, ARHGEF4, EMX1, SP9, CLEC11A, ST8SIA1, BMP3, KCNA3, DMRTA2, KCNK12, CD1D, PRKCB, CYP26C1, ZNF568, ABCB1, ELOVL2, PKIA, SFMBT2 (893), PCBP3, MATK, GRN2D, NDRG4, DLX4, PPP2R5C, FGF14, ZNF132, CHST2 (7890), FLI1, c13orf18, or ZNF569, and

- 5 2) detecting gastric 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

10 following steps:

- 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 15 least one marker selected from a chromosomal region having an annotation selected from the group consisting of ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1, and 2) detecting gastric cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

20

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 100%, or from about 80% to about 90%, or from about 80% to about 85%.

25

Genomic DNA may be isolated by any means, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants, e.g., by digestion with proteinase K. The genomic DNA is then recovered from the solution. This 30 may be carried out by means of a variety of methods including salting out, organic extraction, or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense, and required quantity of DNA. All clinical sample types comprising neoplastic matter or pre-neoplastic matter are suitable for use in the present method, e.g., cell lines, histological slides, biopsies, paraffin-embedded tissue, body fluids,

stool, gastric 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 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 5 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 1-274, e.g., as provided by Tables 10 1, 2, 4 and 7).

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.

15 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 bisulfate (hydrogen sulfite, disulfite) followed by alkaline hydrolysis.

20 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 1-274, e.g., as provided in Tables 1, 2, 4 and 7). 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.

25 Aberrant methylation, more specifically hypermethylation of a marker comprising a DMR (e.g., DMR 1-274, e.g., as provided by Tables 1, 2, 4 and 7) is associated with a gastric cancer and, in some embodiments, predicts tumor site.

The technology relates to the analysis of any sample associated with a gastric cancer. For example, in some embodiments the sample comprises a tissue and/or biological fluid 30 obtained from a patient. In some embodiments, the sample comprises a secretion. In some embodiments, the sample comprises blood, serum, plasma, gastric secretions, pancreatic juice, a gastrointestinal biopsy sample, microdissected cells from a gastrointestinal biopsy, gastrointestinal cells sloughed into the gastrointestinal lumen, and/or gastrointestinal cells recovered from stool. In some embodiments, the subject is human. These samples may

originate from the upper gastrointestinal tract, the lower gastrointestinal tract, or comprise cells, tissues, and/or secretions from both the upper gastrointestinal tract and the lower gastrointestinal tract. The sample may include cells, secretions, or tissues from the liver, bile ducts, pancreas, stomach, colon, rectum, esophagus, small intestine, appendix, duodenum, 5 polyps, gall bladder, anus, and/or peritoneum. In some embodiments, the sample comprises cellular fluid, ascites, urine, feces, pancreatic fluid, fluid obtained during endoscopy, blood, mucus, or saliva. In some embodiments, the sample is a stool sample.

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, 10 while blood, ascites, serum, or pancreatic fluid samples can be obtained parenterally by using a needle and syringe, for instance. Cell free or substantially cell free samples can be obtained by subjecting the sample to various techniques known to those of skill in the art which include, but are not limited to, centrifugation and filtration. Although it is generally preferred that no invasive techniques are used to obtain the sample, it still may be preferable to obtain 15 samples such as tissue homogenates, tissue sections, and biopsy specimens

In some embodiments, the technology relates to a method for treating a patient (e.g., a patient with gastric cancer (e.g., stomach cancer), with early stage gastric cancer, or who may develop gastric 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 20 results of determining the methylation state. The treatment may be administration of a pharmaceutical compound, a vaccine, performing a surgery, imaging the patient, performing another test. Preferably, said use is in a method of clinical screening, a method of prognosis assessment, a method of monitoring the results of therapy, a method to identify patients most likely to respond to a particular therapeutic treatment, a method of imaging a patient or 25 subject, and a method for drug screening and development.

In some embodiments of the technology, a method for diagnosing a gastric cancer (e.g., stomach cancer) in a subject is provided. The terms “diagnosing” and “diagnosis” as used herein refer to methods by which the skilled artisan can estimate and even determine whether or not a subject is suffering from a given disease or condition or may develop a 30 given disease or condition in the future. The skilled artisan often makes a diagnosis on the basis of one or more diagnostic indicators, such as for example a biomarker (e.g., a DMR as disclosed herein), the methylation state of which is indicative of the presence, severity, or absence of the condition.

Along with diagnosis, clinical cancer prognosis relates to determining the aggressiveness of the cancer and the likelihood of tumor recurrence to plan the most effective therapy. If a more accurate prognosis can be made or even a potential risk for developing the cancer can be assessed, appropriate therapy, and in some instances less severe therapy for the 5 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 10 making 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 15 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 gastric 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 20 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 25 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 30 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 gastric cancer risk, prognosis, determining treatment efficacy, and/or progression of the cancer in the subject.

5 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 10 prognostic biomarkers can be made, and a temporal change in the marker can be used to determine a diagnosis or prognosis. For example, a diagnostic marker can be determined at an initial time, and again at a second time. In such embodiments, an increase in the marker from the initial time to the second time can be diagnostic of a particular type or severity of cancer, or a given prognosis. Likewise, a decrease in the marker from the initial time to the second 15 time can be indicative of a particular type or severity of cancer, or a given prognosis.

Furthermore, the degree of change of one or more markers can be related to the severity of the cancer and future adverse events. The skilled artisan will understand that, while in certain embodiments comparative measurements can be made of the same biomarker at multiple time points, one can also measure a given biomarker at one time point, and a second biomarker at 20 a second time point, and a comparison of these markers can provide diagnostic information.

As used herein, the phrase “determining the prognosis” refers to methods by which the skilled artisan can predict the course or outcome of a condition in a subject. The term “prognosis” does not refer to the ability to predict the course or outcome of a condition with 100% accuracy, or even that a given course or outcome is predictably more or less likely to 25 occur based on the methylation state of a biomarker (e.g., a DMR). Instead, the skilled artisan will understand that the term “prognosis” refers to an increased probability that a certain course or outcome will occur; that is, that a course or outcome is more likely to occur in a subject exhibiting a given condition, when compared to those individuals not exhibiting the condition. For example, in individuals not exhibiting the condition (e.g., having a normal 30 methylation state of one or more DMR), the chance of a given outcome (e.g., suffering from a gastric 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 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

5 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%,  
10 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  
15 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  
20 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.

25 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.,  
30 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 gastric cancer. In certain embodiments of the method, a subject is identified as having metaplasia upon identifying an

aberrant methylation state of one or more DMR provided herein in a biological sample obtained from the subject. In other embodiments of the method, the detection of an aberrant methylation state of one or more of such biomarkers in a biological sample obtained from the subject results in the subject being identified as having cancer.

5 The analysis of markers can be carried out separately or simultaneously with additional markers within one test sample. For example, several markers can be combined into one test for efficient processing of a multiple of samples and for potentially providing greater diagnostic and/or prognostic accuracy. In addition, one skilled in the art would recognize the value of testing multiple samples (for example, at successive time points) from 10 the same subject. Such testing of serial samples can allow the identification of changes in marker methylation states over time. Changes in methylation state, as well as the absence of change in methylation state, can provide useful information about the disease status that includes, but is not limited to, identifying the approximate time from onset of the event, the presence and amount of salvageable tissue, the appropriateness of drug therapies, the 15 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 20 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 gastric cancer if, when compared to a control methylation state, there is a measurable difference in the methylation state of at least one biomarker in the sample. Conversely, when no change in methylation 25 state is identified in the biological sample, the subject can be identified as not having gastric 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 gastric cancer can be placed on a more intensive and/or regular screening schedule, including 30 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 gastric cancer has appeared in those subjects.

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

Further with respect to diagnostic methods, a preferred subject is a vertebrate subject. A preferred vertebrate is warm-blooded; a preferred warm-blooded vertebrate is a mammal. A preferred mammal is most preferably a human. As used herein, the term "subject" includes both human and animal subjects. Thus, veterinary therapeutic uses are provided herein. As such, the present technology provides for the diagnosis of mammals such as humans, as well as those mammals of importance due to being endangered, such as Siberian tigers; of economic importance, such as animals raised on farms for consumption by humans; and/or animals of social importance to humans, such as animals kept as pets or in zoos. Examples of such animals include but are not limited to: carnivores such as cats and dogs; swine, including pigs, hogs, and wild boars; ruminants and/or ungulates such as cattle, oxen, sheep, giraffes, deer, goats, bison, and camels; and horses. Thus, also provided is the diagnosis and treatment of livestock, including, but not limited to, domesticated swine, ruminants, ungulates, horses (including race horses), and the like. The presently-disclosed subject matter further includes a system for diagnosing a gastric cancer (e.g., stomach 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 gastric cancer or diagnose a gastric 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, 2, 4 and 7.

## EXAMPLES

### Example 1 – Identifying markers for Detection of Stomach Cancer

Experiments conducted during the course of developing embodiments for the present technology identified 123 DNA methylation markers corresponding to 248 DMRs (e.g.,

5 DMR Nos. 1-248; Table 1) for stomach cancer. Such DNA methylation markers were identified from data generated by CpG island enrichment coupled with massively parallel sequencing of a case - control tissue sample set. Controls included normal stomach tissue, normal colonic epithelia, and normal white blood cell derived DNA. The technique utilized reduced representation bisulfite sequencing (RRBS). The tertiary step involved parsing the  
10 data in terms of coverage cutoffs, excluding all non-informative sites, contrasting the percentage methylation between diagnostic subgroups using logistic regression, creating in-silico CpG “islands” based on defined groupings of contiguous methylation sites, and  
15 Receiver Operating Characteristic analysis. The quaternary step filtered the % methylation data (both individual CpGs and the in-silico clusters) to select markers which maximize signal to noise ratios, minimize background methylation, account for tumor heterogeneity, and to emphasize ROC performance. Furthermore, this analytic approach insured  
identification of the hotspot CpGs over a defined DNA length for easy and optimal design of downstream marker assays - methylation specific PCR, small fragment deep sequencing, etc.

Each sample yielded approximately 2-3 million high quality CpGs, which after  
20 analysis and filtering resulted in less than 1000 highly discriminate individual sites. These clustered into 123 localized regions of differential methylation, some extending for 30-40 bases and some greater than a kilobase (see, Table 1). 12 DMRs were identified with no annotation and no references anywhere in the literature. Such DMRs were named MAX, followed by chromosomal location and coordinates (see, Table 1). All DMRs had AUCs of  
25 0.8 or more (some demonstrated perfect discrimination of cases from controls with an AUC of 1.0). In addition to meeting the AUC threshold of 0.8, identified gastric cancer markers had to exhibit >20-fold higher methylation density in tumor than in normal gastric or colonic mucosa and < 1.0 % methylation in non-neoplastic gastric or colonic mucosa.

30 Table 1.

DMR No.	Gene	Chromosome	Island Range	Strand
1	ABCB1	7	87229775-87229856	-
2	ADAM19	5	157002684-157002766	-
3	ADAM23	2	207307964-207308190	+

4	ADAM23	2	207308771-207308890	+
5	ADCY1	7	45613877-45613977	+
6	ADRA1B	5	159343458-159343934	+
7	AFF3	2	100721738-100721967	-
8	AGBL4	1	49242089-49242514	-
9	ALOX5	10	45914474-45914484	+
10	ANK1	8	41754327-41754528	-
11	ANKRD34B	5	79866019-79866185	-
12	ANO8	19	17439445-17439845	-
13	APOBEC3D	22	39417451-39417576	+
14	ARHGEF7	13	111767862-111767954	+
15	BACH2	6	91004976-91005235	-
16	BARX1	9	96715112-96715262	-
17	BARX1	9	96715374-96715603	-
18	BARX1	9	96721501-96721818	-
19	BCL2	18	60985561-60985600	-
20	BCL2	18	60985788-60985828	-
21	BHLHE22	8	65493937-65494105	+
22	BMP3	4	81952135-81952303	+
23	BMP3	4	81952407-81952528	+
24	BMP6	6	7727566-7727907	+
25	BMP8A	1	39980533-39980619	+
26	C12orf39	12	21680721-21680828	+
27	C12orf42	12	103889269-103889591	-
28	C17orf46	17	43339242-43339498	-
29	C1orf70	1	1475560-1475650	-
30	C1orf70	1	1476065-1476127	-
31	C1orf95	1	226737123-226737246	+
32	C1QL3	10	16563608-16563892	-
33	C9orf50	9	132382706-132382909	-
34	CABIN1	22	24552655-24552724	+
35	CACNA1C	12	2162358-2162804	+
36	CACNA1H	16	1202428-1202560	+
37	CBFA2T3	16	89007265-89007441	-
38	CBLN1	16	49315385-49315725	-
39	CBLN1	16	49316162-49316268	-
40	CCDC106	19	56159280-56159589	+
41	CCDC42B	12	113592232-113592553	+
42	CCND2	12	4378158-4378413	+
43	CD1D	1	158150624-158151205	+
44	CDHR1	10	85954203-85954476	+
45	CELF2	10	11059947-11060085	+
46	CHST2	3	142838645-142839023	+
47	CHST2	3	142839223-142839495	+
48	CHST2	3	142839761-142840276	+
49	CLEC11A	19	51228264-51228394	+
50	CLEC11A	19	51228512-51228732	+
51	CLSTN2	3	139654045-139654299	+
52	CNR1	6	88876414-88876452	-
53	COMP	19	18901578-18901783	-

54	CPE	4	166300108-166300239	+
55	CPNE8	12	39299231-39299461	-
56	DAB2IP	9	124461296-124461420	+
57	DBX1	11	20177747-20178529	-
58	DOCK10	2	225906664-225906922	-
59	DOCK10	2	225907226-225907322	-
60	DOCK10	2	225907515-225907632	-
61	DOK1	2	74782380-74782580	+
62	DYNLRB2	16	80574542-80574687	+
63	DYNLRB2	16	80574797-80574927	+
64	EFNB3	17	7608691-7608815	+
65	EIF4E3	3	71803220-71804279	-
66	ELMO1	7	37487867-37488498	-
67	ELOVL2	6	11044395-11044704	-
68	ELOVL5	6	53213597-53214212	-
69	EPB41L3	18	5542876-5542973	-
70	EYA4	6	133562791-133562878	+
71	EYA4	6	133563343-133563397	+
72	FAIM2	12	50297863-50297988	-
73	FAM78A	9	134152466-134152692	-
74	FBN2	5	127874257-127874290	-
75	FER1L4	20	34189084-34189185	-
76	FER1L4	20	34189488-34189566	-
77	FGF2	4	123748398-123748881	+
78	FLT3	13	28674537-28674770	-
79	FMN2	1	240255171-240255264	+
80	FMNL1	17	43298726-43298819	+
81	FOXD4L1	2	114256519-114257141	+
82	FRMD4A	10	13933638-13934256	-
83	GJC1	17	42907705-42907827	-
84	GLB1L3	11	134146132-134146380	+
85	GNG7	19	2702502-2702568	-
86	GPR135	14	59930725-59931142	-
87	GPX7	1	53067933-53068182	+
88	GRASP	12	52400919-52401166	+
89	GRM8	7	126891909-126892517	-
90	GSG1L	16	28074472-28074761	-
91	GSG1L	16	28075145-28075290	-
92	GUCY1A2	11	106888403-106888538	-
93	HECW2	2	197457425-197458011	-
94	HOPX	4	57521377-57521658	-
95	HOPX	4	57521822-57522182	-
96	HOXD8	2	176994098-176994972	+
97	IKZF1	7	50343838-50344453	+
98	INSM1	20	20348122-20348415	+
99	IRF4	6	391980-392182	+
100	ITGA4	2	182321830-182321983	+
101	KCNK12	2	47797187-47797452	-
102	KCNMA1	10	79397844-79397945	-
103	KCNQ3	8	133492336-133493033	-

104	KDM2B	12	122016688-122017150	-
105	KIF5C	2	149633046-149633131	+
106	KLHL29	2	23609989-23610119	+
107	LBH	2	30455594-30455655	+
108	LINGO3	19	2290273-2290393	-
109	LIPJ	10	90342642-90342873	+
110	LOC285548	4	13549245-13549556	-
111	LOC348840	3	197807208-197807313	-
112	LOC440461	17	66195303-66195773	+
113	LOC645323	5	87970260-87970894	-
114	LRRC10B	11	61276493-61276910	+
115	LRRFIP1	2	238600049-238600078	+
116	LY86-AS1	6	6546450-6546598	-
117	MARK1	1	220702225-220702345	+
118	MATK	19	3785828-3786371	-
119	MAX.chr1.110626388-110626787	1	110626388-110626787	-
120	MAX.chr1.244012787-244012875	1	244012787-244012875	-
121	MAX.chr10.22624234-22624571	10	22624234-22624571	-
122	MAX.chr11.120435350-120435981	11	120435350-120435981	-
123	MAX.chr11.123301058-123301255	11	123301058-123301255	-
124	MAX.chr11.123301366-123301506	11	123301366-123301506	-
125	MAX.chr12.133484966-133485857	12	133484966-133485857	-
126	MAX.chr12.4273880-4274012	12	4273880-4274012	-
127	MAX.chr14.37116653-37117410	14	37116653-37117410	-
128	MAX.chr15.40616377-40616419	15	40616377-40616419	-
129	MAX.chr15.89942727-89942918	15	89942727-89942918	-
130	MAX.chr18.5630121-5630725	18	5630121-5630725	-
131	MAX.chr19.22034799-22034887	19	22034799-22034887	-
132	MAX.chr2.118981853-118981946	2	118981853-118981946	-
133	MAX.chr2.119067940-119068174	2	119067940-119068174	-
134	MAX.chr2.177503038-177503122	2	177503038-177503122	-
135	MAX.chr20.1784209-1784461	20	1784209-1784461	-
136	MAX.chr3.128336893-128337084	3	128336893-128337084	-
137	MAX.chr4.113626838-113626942	4	113626838-113626942	-
138	MAX.chr4.113627402-113627507	4	113627402-113627507	-
139	MAX.chr5.42951300-42951769	5	42951300-42951769	-
140	MAX.chr5.42992128-42992212	5	42992128-42992212	-
141	MAX.chr5.42992322-42992768	5	42992322-42992768	-
142	MAX.chr5.42995102-42995181	5	42995102-42995181	-
143	MAX.chr7.121950350-121951056	7	121950350-121951056	-
144	MAX.chr7.25896389-25896501	7	25896389-25896501	-

145	MAX.chr9.113945-114435	9	113945-114435	-
146	MAX.chr9.79627082-79627175	9	79627082-79627175	-
147	MAX.chr9.99983730-99984118	9	99983730-99984118	-
148	MEF2C	5	88185490-88185576	-
149	MIR129-2	11	43601090-43601396	+
150	MMP23B	1	1567269-1567478	+
151	MPZ	1	161275561-161275653	-
152	MPZ	1	161275766-161275996	-
153	MZB1	5	138728509-138729109	-
154	NDRG4	16	58497251-58497370	+
155	NDRG4	16	58497979-58498250	+
156	NEU1	6	31830234-31830331	-
157	NEU1	6	31830456-31830522	-
158	NEU1	6	31830682-31830917	-
159	NEUROG2	4	113437639-113437953	-
160	NFATC4	14	24837540-24838009	+
161	NID2	14	52535758-52536437	-
162	NRXN2	11	64480878-64481067	-
163	OPLAH	8	145106742-145106921	-
164	OSR2	8	99952116-99952427	+
165	OSR2	8	99952591-99953023	+
166	PAX5	9	37027013-37027670	-
167	PDE8B	5	76506245-76506460	+
168	PEBP4	8	22722910-22722994	-
169	PHF21B	22	45405722-45405819	-
170	PLEKHO1	1	150122907-150123157	+
171	PLXNC1	12	94543138-94543232	+
172	PNMAL2	19	46996516-46996841	-
173	POMC	2	25391168-25391616	-
174	POU3F1	1	38510915-38511213	-
175	POU4F1	13	79176470-79176567	-
176	POU4F1	13	79176954-79177042	-
177	POU4F1	13	79177364-79177872	-
178	PPP2R5C	14	102247537-102247929	+
179	PRDM13	6	100051231-100051668	+
180	PRKCB	16	23847288-23847339	+
181	PRKCB	16	23847575-23847699	+
182	PRKCB	16	23847825-23848020	+
183	PRR16	5	119799921-119800390	+
184	PRR5L	11	36399221-36399473	+
185	PYGL	14	51410872-51411219	-
186	RASA3	13	114897059-114897135	-
187	RASSF2	20	4803267-4803703	-
188	RBPMS2	15	65067593-65067619	-
189	RNF165	18	43913775-43914353	+
190	RPRM	2	154334604-154335314	-
191	RPSAP58	19	23941384-23941670	+
192	RSPO3	6	127440413-127441057	+
193	RUNX3	1	25256650-25257322	-
194	SEPTIN9	17	75368800-75369056	+

195	SEPTIN9	17	75369224-75369327	+
196	SFMBT2	10	7450406-7450831	-
197	SFMBT2	10	7451097-7451185	-
198	SFMBT2	10	7452358-7452452	-
199	SFMBT2	10	7452746-7452779	-
200	SFMBT2	10	7453455-7453511	-
201	SLC35F1	6	118228394-118228979	+
202	SLCO4C1	5	101631546-101631741	-
203	SLCO4C1	5	101632152-101632237	-
204	SND1	7	127671993-127672310	+
205	SNX32	11	65601248-65601514	+
206	SP9	2	175201823-175202389	+
207	ST8SIA4	5	100236601-100236784	-
208	ST8SIA6	10	17496306-17496417	-
209	ST8SIA6	10	17496572-17496711	-
210	TBC1D30	12	65218335-65218778	+
211	TBC1D30	12	65220152-65220299	+
212	TCF24	8	67874670-67875083	-
213	TIAM1	21	32930248-32930322	-
214	TLX2	2	74742033-74742264	+
215	TM6SF1	15	83776207-83776706	+
216	TRIM9	14	51560661-51561240	-
217	TRPC3	4	122872067-122872455	-
218	TRPC3	4	122872626-122872724	-
219	TSC22D4	7	100075304-100075445	-
220	TTC28	22	29075259-29075493	-
221	UNC5CL	6	40995803-40996093	-
222	VCAN	5	82768837-82769031	+
223	VEGFC	4	177713711-177713910	-
224	WNT5A	3	55522631-55522773	-
225	WT1	11	32457189-32457408	-
226	ZBTB16	11	113930478-113930929	+
227	ZDBF2	2	207139497-207139592	+
228	ZNF132	19	58951402-58951530	-
229	ZNF134	19	58125597-58125764	+
230	ZNF256	19	58459002-58459219	-
231	ZNF331	19	54024023-54024436	+
232	ZNF347	19	53662191-53662618	-
233	ZNF354C	5	178487123-178487607	+
234	ZNF43	19	22018452-22018639	-
235	ZNF43	19	22018852-22019004	-
236	ZNF461	19	37157822-37158261	-
237	ZNF569	19	37957760-37958046	-
238	ZNF570	19	37960066-37960542	+
239	ZNF610	19	52839503-52840013	+
240	ZNF625	19	12267378-12267677	-
241	ZNF665	19	53696101-53696195	-
242	ZNF665	19	53696497-53696668	-
243	ZNF682	19	20149796-20149923	-
244	ZNF71	19	57106756-57106967	+

245	ZNF788	19	12203466-12203641	+
246	ZNF813	19	53970869-53971093	+
247	ZNF813	19	53971239-53971374	+
248	ZNF83	19	53193211-53193303	-

22 of the 123 disclosed markers were chosen based on AUC and fold change ratios (cancer vs. normal stomach + normal colon) and taken into a larger tissue validation study  
5 with 73 discovery markers from other GI cancer sites (colon, pancreas, bile duct, esophagus). Samples in these different cohorts ranged between 15 and 50 samples, both cancer, normal, and if applicable adenomas. The validation platform was quantitative methylation specific PCR with primers optimized on methylation control sets. All assays were performed on Roche 480 LightCyclers. Results were analyzed for ROC characteristics, fold change, and  
10 complementarity. The top 30, ranked by sensitivity (for both cancer and adenoma) are provided in Table 2. Table 3 provides the primer information for the markers shown in Table 2. Of these, 10 were chosen for further tissue validation in a joint Mayo Korean study (see, Example 2).

15 Table 2

DMR No.	Gene	sensitivity (100% spec)	Chromosome	Island Range
253	ELMO1	95%	7	37487755-37488477
251	ARHGEF4	91%	2	131797843-131797938
254	EMX1	88%	2	73147710-73147772
255	SP9	86%	2	175202025-175202158
256	CLEC11A	86%	19	51228217-51228732
249	ST8SIA1	86%	12	22487528-22487620
257	BMP3	86%	4	81952348-81952402
258	KCNA3	84%	1	111217654-111217816
259	DMRTA2	81%	1	50884349-50884499
260	KCNK12	81%	2	47797260-47797371
261	CD1D	79%	1	158150797-158151205
262	PRKCB	79%	16	23847825-23848025
250	CYP26C1	79%	10	94822416-94822607
263	ZNF568	79%	19	37407197-37407365
1	ABCB1	79%	7	87229775-87229856

264	ELOVL2	77%	6	11044395-11044834
265	PKIA	77%	8	79428485-79428684
196	SFMBT2 (893)	74%	10	7450406-7450831
266	PCBP3	72%	21	47063793-47064177
118	MATK	72%	19	3785828-3786371
267	GRIN2D	70%	19	48918144-48918350
268	NDRG4	70%	16	58497395-58497451
269	DLX4	70%	17	48042426-48042820
270	PPP2R5C	67%	14	102247525- 102247929
271	FGF14	67%	13	103046885- 103047269
272	ZNF132	67%	19	58951402-58951775
46	CHST2 (7890)	65%	3	142838645- 142839023
273	FLI1	63%	11	128563956- 128564209
252	c13orf18	63%	13	46960767-46961669
237	ZNF569	70% (92% spec)	19	37957760- 37958046

Table 3.

DMR No.	Gene	Forward MSP Primer (5'-3')	Reverse MSP Primer (5'-3')
253	ELMO1	TTA TAT TTT TCG TTT TTA GTA ATT TCG CGT TAG C (SEQ ID NO: 21)	GAA AAC CCG CCG AAA CAT TTC GA (SEQ ID NO: 22)
251	ARHGEF4	TGT TTT CGC GGT CGT TAT ATA TTA CGT CGT (SEQ ID NO: 23)	GAA CTA TCC CCG AAC TCC GAC TCG A (SEQ ID NO: 24)
254	EMX1	CGG GTT TTA GCG ATG TTT ATT TTA GTT TCG T (SEQ ID NO: 25)	CCT TTT CGT TCG TAT AAA ATT TCG TT (SEQ ID NO: 26)
255	SP9	CGG GGG TAA AAA GGG TAG CGA TAG TGA TAC (SEQ ID NO: 27)	CGA AAA ATT CGA AAC GAA ACG TC (SEQ ID NO: 28)
256	CLEC11A	AGT TTG GCG TAG TCG GTA GAT C (SEQ ID NO: 29)	GCG CGC AAA TAC CGA ATA AAC G (SEQ ID NO: 30)
249	ST8SIA1	GAC GTT TGT CGT CGG GTT CGT TC (SEQ ID NO: 13)	AAA AAC CCT CCG CTA CCA CTT CGC (SEQ ID NO: 14)
257	BMP3	GTTTAATTTCGGTTTCGTCGTC (SEQ ID NO: 31)	CGCTACGAAACACTCCGA (SEQ ID NO: 32)
258	KCNA3	TTT TGT TGG GTG GGT ATT TTC GCG (SEQ ID NO: 33)	ACA ACC CTC AAA ACC CCT CGA TAT C (SEQ ID NO: 34)
259	DMRTA2	TTT ATT TAT GGA TTA CGT TTT TAG CGA (SEQ ID NO: 35)	GAC GAC CGA ACG ATC ACG CA (SEQ ID NO: 36)
260	KCNK12	GCG TCG TTA GTA GTA CGA AGC (SEQ ID NO: 37)	GCA CCT CAA CGA AAA CAC CGA (SEQ ID NO: 38)

261	CD1D	CGC CGT AGC GGC GTT TC (SEQ ID NO: 39)	CCC ATA TCG CCC GAC GTA A (SEQ ID NO: 40)
262	PRKCB	GTC GTT TTT GGG CGT TTT AGA GGC (SEQ ID NO: 41)	TAATCCCAAACGAAAC CGC CG (SEQ ID NO: 42)
250	CYP26C1	GGT TTT TTG GTT ATT TCG GAA TCG T (SEQ ID NO: 43)	TAT AAA AAC GCG CGT AAT CAA CGC T (SEQ ID NO: 44)
263	ZNF568	TTG AGA TGT TGG GTG AAG GCG ATT C (SEQ ID NO: 45)	CGC TAA CGC GAA AAA ATA ATT CGA CG (SEQ ID NO: 46)
1	ABCB1	GAT TTT GTT CGT CGT TAG TGC (SEQ ID NO: 9)	TCT CTA AAC CCG CGA ACG A (SEQ ID NO: 10)
264	ELOVL2	CGG TTT TAT TTA TTA TGA TTC GTA GCG G (SEQ ID NO: 47)	CGA CTA CCC TAA ACA ACG CAT CGC (SEQ ID NO: 48)
265	PKIA	CGG GGA TGA TTT TAT GTA GTC GGA GTT TCG C (SEQ ID NO: 49)	CCC GCC GAA TAC TCG ATC AAC TCG (SEQ ID NO: 50)
196	SFMBT2 (893)	AAA ACG TTT TTA GGT ATT TGG TCG T (SEQ ID NO: 5)	ATA AAA AAA CCT CCT CCG AAC CCG C (SEQ ID NO: 6)
266	PCBP3	GGT CGC GTC GTT TTC GAT C (SEQ ID NO: 51)	GCC GCA AAC GCC GAC GA (SEQ ID NO: 52)
118	MATK	TGT ATA TTT CGA GGC GGT TTC GG (SEQ ID NO: 53)	CGC CCC CAA AAT AAA AAA ACG AA (SEQ ID NO: 54)
267	GRIN2D	TTT TTC GTT CGT TTT TTT ATT TTT TCG A (SEQ ID NO: 55)	AAC GCA TAC CAT CGA CTT CAA CTA CGA C (SEQ ID NO: 56)
268	NDRG4	CGGTTTCGTTCGTTTTTCG (SEQ ID NO: 57)	CCGCCTTCTACCGCGACTA (SEQ ID NO: 58)
269	DLX4	GTA TTT TTA ATA TTT GGT GAG TGC G (SEQ ID NO: 59)	TAC TAA AAC GTA CGA TAA ACA TCG T (SEQ ID NO: 60)
270	PPP2R5C	TCG ATT TTA TTT TTG TTG TCG TTG TAG ATT CGC (SEQ ID NO: 61)	GAA AAA ACT AAA AAA CGA CAA AAA AAC CCG ACG (SEQ ID NO: 62)
271	FGF14	GGA GGG AGT TTA ACG GGT TAG GTA C (SEQ ID NO: 63)	CGT AAA CAA AAA ACA ACG ACG ACG (SEQ ID NO: 64)
272	ZNF132	ATT GCG TCG TTG TTT AGG TAA CGT A (SEQ ID NO: 65)	CGA AAA TAC CTA TCT CCT CGA CT (SEQ ID NO: 66)
46	CHST2 (7890)	GGA ACG AGT GAT AGT CGG ATA GTT CGT C (SEQ ID NO: 67)	CGC CCG AAA ACG ACC CCG (SEQ ID NO: 68)
273	FLI1	GGG AGT GAG GGT AGG GCG TTC (SEQ ID NO: 69)	CTC GCA ACC CCT TCG AAT TAA CCC G (SEQ ID NO: 70)
252	c13orf18	TTT AGG GAA GTA AAG CGT CGT TTT C (SEQ ID NO: 19)	AAC GAC GTC TCG ATA CCT ACG A (SEQ ID NO: 20)
237	ZNF569	TGT GGA ATC GGG GTT TGT GTT CGC (SEQ ID NO: 11)	CCC ACC CAA CAC AAA AAA TCC GAC G (SEQ ID NO: 12)

**Example 2 – Gastric Cancer Detection by Novel Methylated DNA Markers:  
Tissue Validation in Patient Cohorts from the United States and South Korea**

Experiments conducted during the course of developing embodiments for the present technology identified novel methylated DNA markers by whole methylome sequencing and 5 then validated best candidates in tissues from patient cohorts in the U.S. and South Korea (see, Tables 4, 5, and 6).

Following a whole methylome discovery effort using reduced representation bisulfite sequencing, 17 methylated DNA marker candidates were selected for blinded validation in 10 gastric tissue sets from patients in the U.S. and South Korea. DNA from micro-dissected tissues was assayed by methylation-specific polymerase chain reaction; marker levels were 15 standardized with beta-actin (total human DNA). GC cases comprised 35 patients from the U.S. and 50 from Korea with pathologically-confirmed intact adenocarcinoma. Controls included histologically-normal gastric epithelia from 65 healthy U.S. patients demographically matched to cases and from the same 50 Korean case patients but obtained at sites distant from the primary GC. For each marker, area under the ROC curve (AUC) was calculated from nominal logistic regression.

Overall, the top 10 markers included ARHGEF4, ELMO1, ABCB1, CLEC11A, 20 ST8SIA1, SFMBT2, CD1D, CYP26C1, ZNF569, and C13ORF18 (see, Table 4). At 95% specificity, a panel of these markers detected GC in 100% of the U.S. cohort and 94% of the 25 South Korean cohort. Respective individual AUCs in the U.S. cohort were 0.92, 0.91, 0.89, 0.88, 0.85, 0.82, 0.82, 0.81, 0.75, and 0.66 and in the Korean cohort were 0.73, 0.90, 0.78, 0.75, 0.80, 0.94, 0.87, 0.79, 0.75, and 0.79 (see, Table 6). Some markers, like ELMO1, exhibited similarly high discrimination for GC in each patient cohort; other markers, like ARHGEF4, showed more variable discrimination due to high background on normal controls (NL) across patient cohorts (see, Fig. 1). Table 5 provides forward and reverse primer information for the 10 markers shown in Table 4.

Table 4

DMR No.	Gene	chromosome	Island Range
49	CLEC11A	19	51228264-51228394
43	CD1D	1	158150624-158151205
196	SFMBT2 (893)	10	7450406-7450831
66	ELMO1	7	37487867-37488498
1	ABCB1	7	87229775-87229856
237	ZNF569	19	37957760-37958046

249	ST8SIA1	12	22487528-22487620
250	CYP26C1	10	94822416-94822607
251	ARHGEF4	2	131797843-131797938
252	c13orf18	13	46960767-46961669

Table 5

DMR No.	Gene	Forward Primer (5'-3')	Reverse Primer (3'-5')
49	CLEC11A	AGT TTG GCG TAG TCG GTA GAT C (SEQ ID NO: 1)	GCG CGC AAA TAC CGA ATA AAC G (SEQ ID NO: 2)
43	CD1D	GCG CGT AGC GGC GTT TC (SEQ ID NO: 3)	CCC ATA TCG CCC GAC GTA A (SEQ ID NO: 4)
196	SFMBT2 (893)	AAA ACG TTT TTA GGT ATT TGG TCG T (SEQ ID NO: 5)	ATA AAA AAA CCT CCT CCG AAC CCG C (SEQ ID NO: 6)
66	ELMO1	GGA GGG AGT TTA ACG GGT TAG GTA C (SEQ ID NO: 7)	CGT AAA CAA AAA ACA ACG ACG ACG (SEQ ID NO: 8)
1	ABCB1	GAT TTT GTT CGT CGT TAG TGC (SEQ ID NO: 9)	TCT CTA AAC CCG CGA ACG A (SEQ ID NO: 10)
237	ZNF569	TGT GGA ATC GGG GTT TGT GTT CGC (SEQ ID NO: 11)	CCC ACC CAA CAC AAA AAA TCC GAC G (SEQ ID NO: 12)
249	ST8SIA1	GAC GTT TGT CGT CGG GTT CGT TC (SEQ ID NO: 13)	AAA AAC CCT CCG CTA CCA CTT CGC (SEQ ID NO: 14)
250	CYP26C1	GGT TTT TTG GTT ATT TCG GAA TCG T (SEQ ID NO: 15)	TAT AAA AAC GCG CGT AAT CAA CGC T (SEQ ID NO: 16)
251	ARHGEF4	TGT TTT CGC GGT CGT TAT ATA TTA CGT CGT (SEQ ID NO: 17)	GAA CTA TCC CCG AAC TCC GAC TCG A (SEQ ID NO: 18)
252	c13orf18	TTT AGG GAA GTA AAG CGT CGT TTT C (SEQ ID NO: 19)	AAC GAC GTC TCG ATA CCT ACG A (SEQ ID NO: 20)

Table 6

DMR No.	Gene	Area Under Curve US Cohort	Area Under Curve Korean Cohort
49	CLEC11A	0.88	0.75
43	CD1D	0.82	0.87
196	SFMBT2 (893)	0.82	0.94
66	ELMO1	0.91	0.9
1	ABCB1	0.89	0.78
237	ZNF569	0.75	0.75
249	ST8SIA1	0.85	0.8
250	CYP26C1	0.81	0.79

251	ARHGEF4	0.92	0.73
252	c13orf18	0.66	0.79

**Example 3 – Detection of Gastric Cancer by Assay of Novel Methylated DNA Markers in Plasma**

This example explored the use of selected methylated DNA marker (MDM) candidates from Examples I and II to assay from plasma as an approach to gastric adenocarcinoma (GC) detection. Indeed, this example demonstrates that a panel of novel MDMs assayed in plasma accurately discriminates GC cases from normal controls. It was shown that marker levels are negligible in controls but elevated in cases and increase progressively with GC stage.

Archival plasma samples that met case entry criteria from a single center were included in the study. Cases comprised 37 patients with pathologically-confirmed intact adenocarcinoma across all stages, histologic types, and gastric sites. Case plasma samples had been collected prior to resection, chemotherapy, or radiation. As controls, archival plasma samples from 38 age and sex-matched healthy volunteers were selected. DNA was extracted from 2 ml plasma and bisulfite treatment. Selected MDMs (ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1) (Table 7 provides DMR information for LRCC4) were then assayed using a non-optimized quantitative allele-specific real-time target and signal amplification method. These markers had demonstrated high performance in tissue ranked by clinical sensitivity and specificity, fold change in methylation signal when comparing gastric cancer vs. normal gastric tissues, and very low background methylation in blood cell derived DNA.

To facilitate the highest analytical performance in plasma, these markers - which were previously queried in tissue using methylation specific PCR assays - were re-designed using the original DMR sequences as multiplexed QuARTs (quantitative allele-specific real-time target and signal amplification) assays. QuARTs assays include PCR primers (see, Table 8), detection probes (see, Table 8), and invasive oligos (Integrated DNA Technologies), GoTaq DNA Polymerase (Promega), Cleavase II (Hologic), and fluorescence resonance energy transfer reporter cassettes (FRETs) containing FAM, HEX, and Quasar 670 dyes (Biosearch Technologies).

Fig. 2 provides the oligonucleotide sequences for the FRET cassettes, used in the detection of methylated DNA signatures by QuARTs (quantitative allele-specific real-time

target and signal amplification) assay. Each FRET sequence includes a fluorophore and quencher which can be multiplexed together into 3 separate assays.

Standards were derived from target sequences grown and cut from plasmid controls. Absolute copies of serially diluted standards were obtained from Poisson modeling.

5 DNA was purified from 2mL of archival plasma samples and subsequently bisulfite converted using in-house proprietary methods and automated instrumentation. A process control gene was added up front to all samples to correct for target recovery variability during the study.

10 Samples were preamplified using the 12 primer sets, diluted, and subjected to QuARTs amplification. The platform for the latter step was the LightCycler 480 (Roche). Strand counts from each of the markers were normalized to strands from a  $\beta$ -actin control assay which was included in the multiplex reaction.

15 Results were analyzed logically and plotted in a matrix format to assess complementarity.

Fig. 3 provides receiver operating characteristic curve highlighting the performance in plasma of the 3 marker panel (ELMO1, ZNF569, and c13orf18) compared with individual marker curves. At 100% specificity, the panel detected gastric cancer at a sensitivity of 86%.

20 Fig. 4A provides log scale absolute strand count of methylated ELMO1 in plasma by stage illustrating the progression from normal mucosa to stage 4 gastric cancer. Quantitative marker levels increased directly with GC stage.

Fig. 4B provides a bar chart demonstrating gastric cancer sensitivity in plasma by stage (at 100% specificity) of the 3 marker panel (ELMO1, ZNF569, and c13orf18).

Indeed, ELMO1 was the most discriminant marker and alone had an AUC of 94% (95% CI 89-99%). At 100% specificity, a panel of 3 MDMs (ELMO1, ZNF569, and C13orf18) detected GC at a sensitivity of 86% (71-95%). By GC stage (Fig. 4A), sensitivities by the panel at 100% specificity were 50%, 92%, 100%, and 100%, for stages 1 (n=8), 2 (n=13), 3 (n=4), and 4 (n=11), respectively (p=0.01, for trend). Quantitative MDM levels increased directly with GC stage. As shown with distributions of methylated ELMO1 (Fig. 4B), the median plasma strand count/ml plasma in controls was 0.4 and strand counts in 30 GC cases at stages 1, 2, 3, and 4 were 16, 111, 101, and 213, respectively (p=0.01, for trend). Neither age nor sex affected marker levels.

Fig. 5A, B and C provides performance of 12 gastric cancer markers (ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1; ELMO1 was run twice, the second time in a bplex format). in 74

plasma samples in a matrix format at 90% specificity (A), 95% specificity (B), and 100% specificity (C). Markers are listed vertically and samples horizontally. Samples are arranged with normals on the left and cancers on the right. Positive hits are in light grey and misses in dark grey. Here, the top performing marker – ELMO1 – is listed first at 90% specificity and 5 the remaining panel at 100% specificity. This plot allows markers to be assessed in a complementary fashion.

Fig. 6 provides performance of the 3 gastric cancer markers (final panel) in 74 plasma samples at 100% specificity. Markers are listed vertically and samples horizontally. Samples are arranged with normals on the left and cancers on the right. The cancers are arranged by 10 stage. Positive hits are in light grey and misses in dark grey (ELMO1, ZNF569, and c13orf18).

Fig. 7A-M provides box plots (linear scale) of 12 gastric cancer markers in plasma (ELMO1, ZNF569, C13orf18, CD1D, ARHGEF4, SFMBT2, PPP25RC, CYP26C1, PKIA, CLEC11A, LRRC4, and ST8SIA1; ELMO1 was run twice, the second time in a bplex 15 format). Control samples (N=38) are listed on the left and gastric cancer cases (N=36) on the right. The vertical axis is % methylation normalized to  $\beta$ -actin strands.

Table 7.

DMR No.	Gene	chromosome	Island Range
274	LRCC4	7	127671993-127672310

20 Table 8.

MARKER	OLIGO TYPE	OLIGO NAME	SEQUENCE
PKIA	Primer	F Primer PKIA	GCGAGTTGATCGAGTATTGG (SEQ ID NO: 71)
	Primer	R Primer PKIA	CTAATCAACGCAAATCCCCGC (SEQ ID NO: 72)
	Probe	PKIA_Pb_A5	CCACGGACGCGCACGCCGCCA/3C6/ (SEQ ID NO: 73)
C13ORF18	Primer	C13ORF18_FP_1	CGTAGTTTTAGAAGTTTTAGGGA AGT (SEQ ID NO: 74)
	Primer	C13ORF18_RP_1	CGACGTCTCGATACCTACGA (SEQ ID NO: 75)
	Probe	C13ORF18_PROBE_1	CCACGGACGACGAAAACGACG /3C6/ (SEQ ID NO: 76)
CLEC11A	Primer	CLEC_FP_1	GCGGGAGTTGGCGTAG (SEQ ID NO: 77)
	Primer	CLEC_RP_1	CGCGCAAATACGAATAAACG (SEQ ID NO: 78)
	Probe	CLEC_PROBE_1	CCACGGACGGTCGGTAGATCG/3C6/

			(SEQ ID NO: 79)
ELMO	Primer	F Primer ELMO1	GTTTCCGGTTTCGGGTTTG (SEQ ID NO: 80)
	Primer	R Primer ELMO1	CGCTAAAAATACGACGCCG (SEQ ID NO: 81)
	Probe	Probe A5 ELMO1 v1a	CCACGGACGGCGTTTATTG /3C6/ (SEQ ID NO: 82)
SFMBT2 (896)	Primer	F Primer SFMBT2_896 ver2	GCGTTTAGGTTGGTCGGAG (SEQ ID NO: 83)
	Primer	R Primer SFMBT2_896 ver2	ACGCACTAACCTACGAAC (SEQ ID NO: 84)
	Probe	Probe A5 SFMBT2_896 ver2	CCACGGACGCCGAAAAACTAC/3C6/ (SEQ ID NO: 85)
ZNF569	Primer	ZNF569 FP_V2	AGAGTTCGCGTTAGAGTTAG (SEQ ID NO: 86)
	Primer	ZNF569 RP_V2	TCGAAACCTATATCCGCGAAC (SEQ ID NO: 87)
	Probe	ZNF569 Pb A5_V2	CCACGGACGCCGTTTTCGA/3C6/ (SEQ ID NO: 88)
BTACT	Primer	ACTB BT FP1	TTTGTTTTTTGATTAGGTGTTAAGA (SEQ ID NO: 89)
	Primer	ACTB BT RP1	CACCAACCTCATAACCTTATC (SEQ ID NO: 90)
	Probe	ACTB BT PB A3	GACGCGGAGAGATAGTGTGTGG /3C6/ (SEQ ID NO: 91)
PPP2R5C	Primer	F Primer PPP2R5C	TTCGATTTATTTTGTGTCGTTGTAGA (SEQ ID NO: 92)
	Primer	PPP2R5C RP_V2	ACGACAAAAAAACCCGACGA (SEQ ID NO: 93)
	Probe	PPP2RC_Pb_A1	CGCCGAGGATTCGCGTGGGT/3C6/ (SEQ ID NO: 94)
ARHGEF 4	Primer	ARHGEF4 FP	CGTCGCGTTATTTATTCGGCG (SEQ ID NO: 95)
	Primer	ARHGEF4 RP	GCTCCTAATTCTCATCACGTCGT (SEQ ID NO: 96)
	Probe	ARHGEF4 Pb A1	CGCCGAGGGCGCGTTTGC/3C6/ (SEQ ID NO: 97)
CD1D	Primer	CD1D_FP_2a	GAAGAGTGCCTAGGTTAGAGG (SEQ ID NO: 98)
	Primer	CD1D_RP_2a	CATATGCCCGACGTAAAAACC (SEQ ID NO: 99)
	Probe	CD1D_PROBE_2a_A1	CGCCGAGGCTCGCAAACGC/3C6/ (SEQ ID NO: 100)
ST8SIA1	Primer	ST8SIA1_FP_ESO	AAATATGTGGTTCGTGGCGTT (SEQ ID NO: 101)
	Primer	ST8SIA1_RP_ESO	ACGCAACAACGCGAAAAAC (SEQ ID NO: 102)
	Probe	ST8SIA1_Pb_A1_ESO	CGCCGAGGCGACGAAAAACG/3C6/ (SEQ ID NO: 103)
CYP26C1	Primer	CYP26C1 FP	CGTCGCGTTATTTATTCGGCG (SEQ ID NO: 104)
	Primer	CYP26C1 RP	GCTCCTAATTCTCATCACGTCGT

			(SEQ ID NO: 105)
LRRC4	Probe	CYP26C1 Pb A1	CGCCGAGGGCGGCCTTTGC/3C6/ (SEQ ID NO: 106)
	Primer	LRRC4 FP	GCGTTAACCGAGGTA (SEQ ID NO: 107)
	Primer	LRRC4 RP	ACAATACTCTATATATTAAACGCCGCT C (SEQ ID NO: 108)
	Probe	LRRC4 Pb A1	CGCCGAGGAGGCGACGGAGG/3C6/ (SEQ ID NO: 109)

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.

Definitions of specific embodiments of the invention as claimed herein follow.

According to a first embodiment of the invention, there is provided a method for characterizing a biological sample comprising:

(a) measuring a methylation level of a CpG site for three-or more genes selected from ELMO1, and at least two further genes selected from ARHGEF4, EMX1, ST8SIA1, CYP26C1, ZNF568, PPP2R5C, and ZNF132 in the biological sample of a human individual through treating genomic DNA in the biological sample with bisulfite;

amplifying the bisulfite-treated genomic DNA using primers specific for each of the selected three or more genes; and

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

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

- (c) determining that the individual has stomach cancer when the methylation level measured in the three or more genes is higher than the methylation level measured in the corresponding control samples.
- 5 According to a further embodiment of the invention, there is provided a method a method for characterizing a biological sample comprising:
- (a) measuring a methylation level of a CpG site for three or more genes selected from ELMO1, and at least two further genes selected from ARHGEF4, EMX1, PPP25RC, CYP26C1, ZNF568m ZNF132 and FLI1 in the biological sample of a human individual through treating genomic DNA in the biological sample with bisulfite;
- amplifying the bisulfite-treated genomic DNA using primers specific for each of the selected three or more genes;
- and
- 5 determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR;
- (b) comparing the methylation level to a methylation level of a corresponding set of genes in control samples without stomach cancer; and
- (c) determining that the individual has stomach cancer when the methylation level measured in the three or more genes is higher than the methylation level measured in the corresponding control samples.

25 In the present specification and claims, the word 'comprising' and its derivatives including 'comprises' and 'comprise' include each of the stated integers but does not exclude the inclusion of one or more further integers.

The reference to any prior art in this specification is not, and should not be taken as an acknowledgement or any form of suggestion that the prior art forms part of the common 30 general knowledge.

## CLAIMS

1. A method for characterizing a biological sample comprising:
  - (a) measuring a methylation level of a CpG site for three or more genes selected from ELMO1, and at least two further genes selected from ARHGEF4, EMX1, ST8SIA1, CYP26C1, ZNF568, PPP2R5C, and ZNF132 in the biological sample of a human individual through treating genomic DNA in the biological sample with bisulfite;
    - amplifying the bisulfite-treated genomic DNA using primers specific for each of the selected three or more genes; and
    - determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR;
  - (b) comparing the methylation level to a methylation level of a corresponding set of genes in control samples without stomach cancer; and
  - (c) determining that the individual has stomach cancer when the methylation level measured in the three or more genes is higher than the methylation level measured in the corresponding control samples.
2. The method of claim 1, wherein the biological sample is a stool sample, a tissue sample, or a plasma sample.
3. The method of claim 2, wherein the tissue sample is gastric tissue.
4. The method of claim 1, wherein said measuring the methylation level a CpG site for three or more genes comprises a determination selected from the group consisting of determining the methylation score of said CpG site and determining the methylation frequency of said CpG site.
5. A method for characterizing a biological sample comprising:
  - (a) measuring a methylation level of a CpG site for three or more genes selected from ELMO1, and at least two further genes selected from ARHGEF4, EMX1, PPP2R5C, CYP26C1, ZNF568, ZNF132, and FLI1 in the biological sample of a

human individual through treating genomic DNA in the biological sample with bisulfite;

amplifying the bisulfite-treated genomic DNA using primers specific for each of the selected three or more genes; and

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

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

(c) determining that the individual has stomach cancer when the methylation level measured in the three or more genes is higher than the methylation level measured in the corresponding control samples.

5 6. The method of claim 5, wherein the biological sample is a stool sample, a tissue sample, or a plasma sample.

7. The method of claim 6, wherein the tissue sample is gastric tissue.

10 8. The method of claim 5, wherein said measuring the methylation level a CpG site for two or more genes comprises a determination selected from the group consisting of determining the methylation score of said CpG site and determining the methylation frequency of said CpG site.

25 9. The method of claim 1, wherein the primers specific for each of the selected three or more genes are selected from:

for ELMO1 a set of primers consisting of SEQ ID NOS: 21 and 22, or SEQ ID NOS: 80 and 81;

30 for ARHGEF4 a set of primers consisting of SEQ ID NOS: 23 and 24, or SEQ ID NOS: 95 and 96;

for EMX1 a set of primers consisting of SEQ ID NOS: 25 and 26,

for ST8SIA1 a set of primers consisting of SEQ ID NOS: 13 and 14, or SEQ ID NOS: 101 and 102;

- for CYP26C1 a set of primers consisting of SEQ ID NOS: 43 and 44, or SEQ ID NOS: 104 and 105;
- for ZNF568 a set of primers consisting of SEQ ID NOS: 45 and 46,
- for PPP2R5C a set of primers consisting of SEQ ID NOS: 61 and 62, or SEQ ID NOS: 92 and 93; and
- for ZNF132 a set of primers consisting of SEQ ID NOS: 65 and 66.
10. The method of claim 5, wherein the primers specific for each of the selected three or more genes are selected from:
- for ELMO1 a set of primers consisting of SEQ ID NOS: 21 and 22, or SEQ ID NOS: 80 and 81;
- for ARHGEF4 a set of primers consisting of SEQ ID NOS: 23 and 24, or SEQ ID NOS: 95 and 96;
- for PPP25RC a set of primers consisting of SEQ ID NOS: 61 and 62, or SEQ ID NOS: 92 and 93;
- for CYP26C1 a set of primers consisting of SEQ ID NOS: 43 and 44, or SEQ ID NOS: 104 and 105;
- for FLI1 a set of primers consisting of SEQ ID NOS: 69 and 70;
- for EMX1 a set of primers consisting of SEQ ID NOS: 25 and 26;
- for ZNF568 a set of primers consisting of SEQ ID NOS: 45 and 46;
- and
- for ZNF132 a set of primers consisting of SEQ ID NOS: 65 and 66.

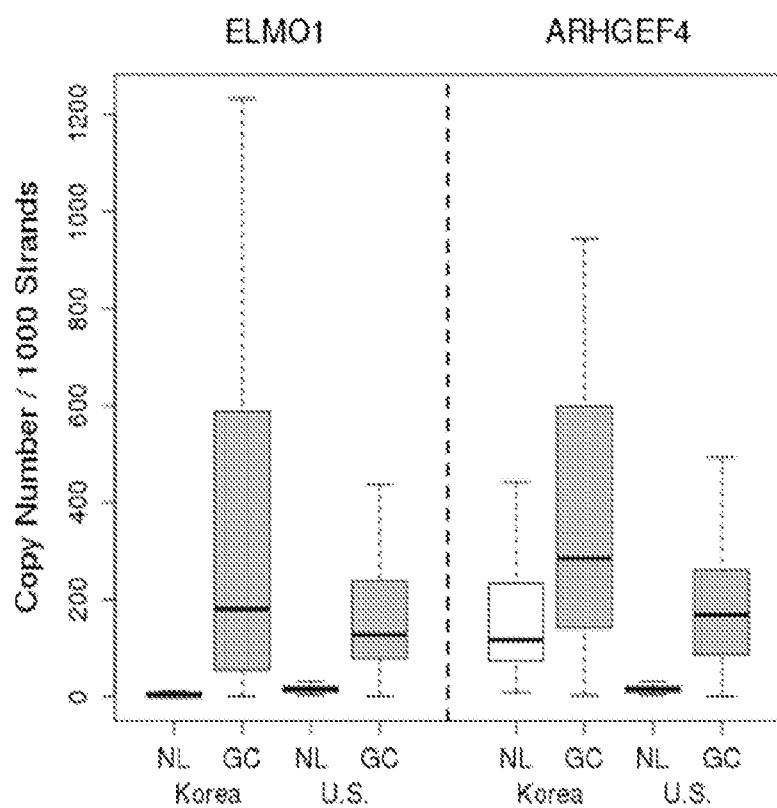
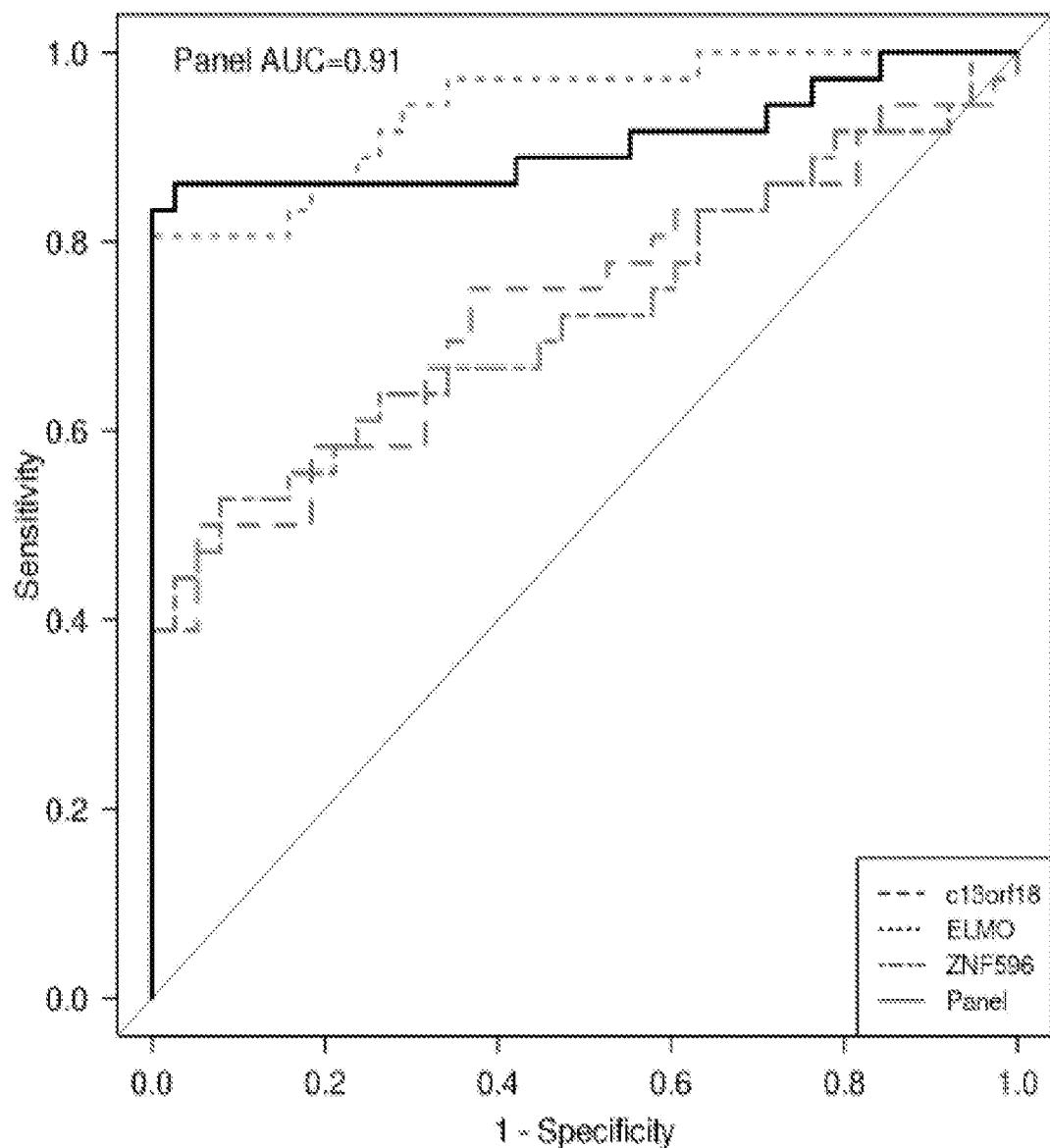
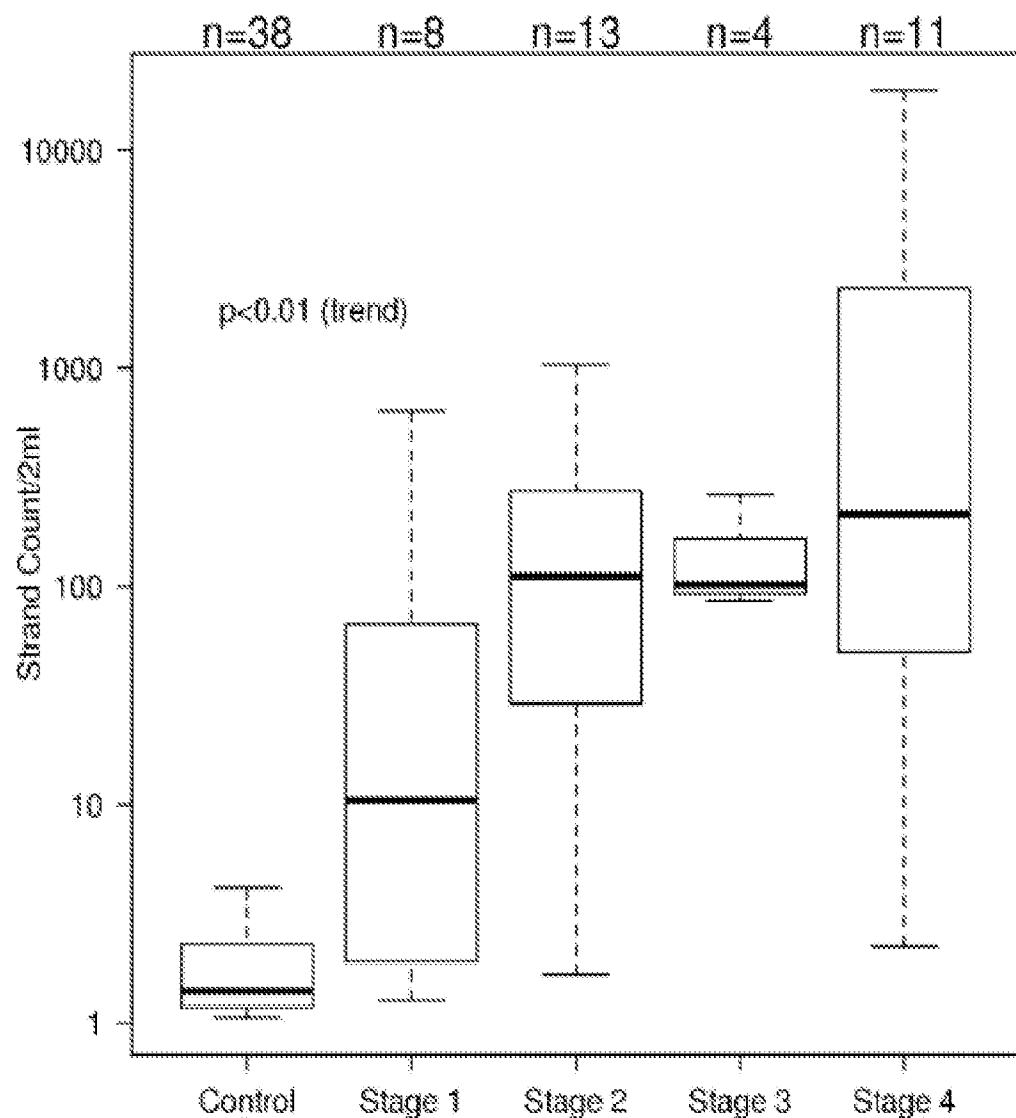
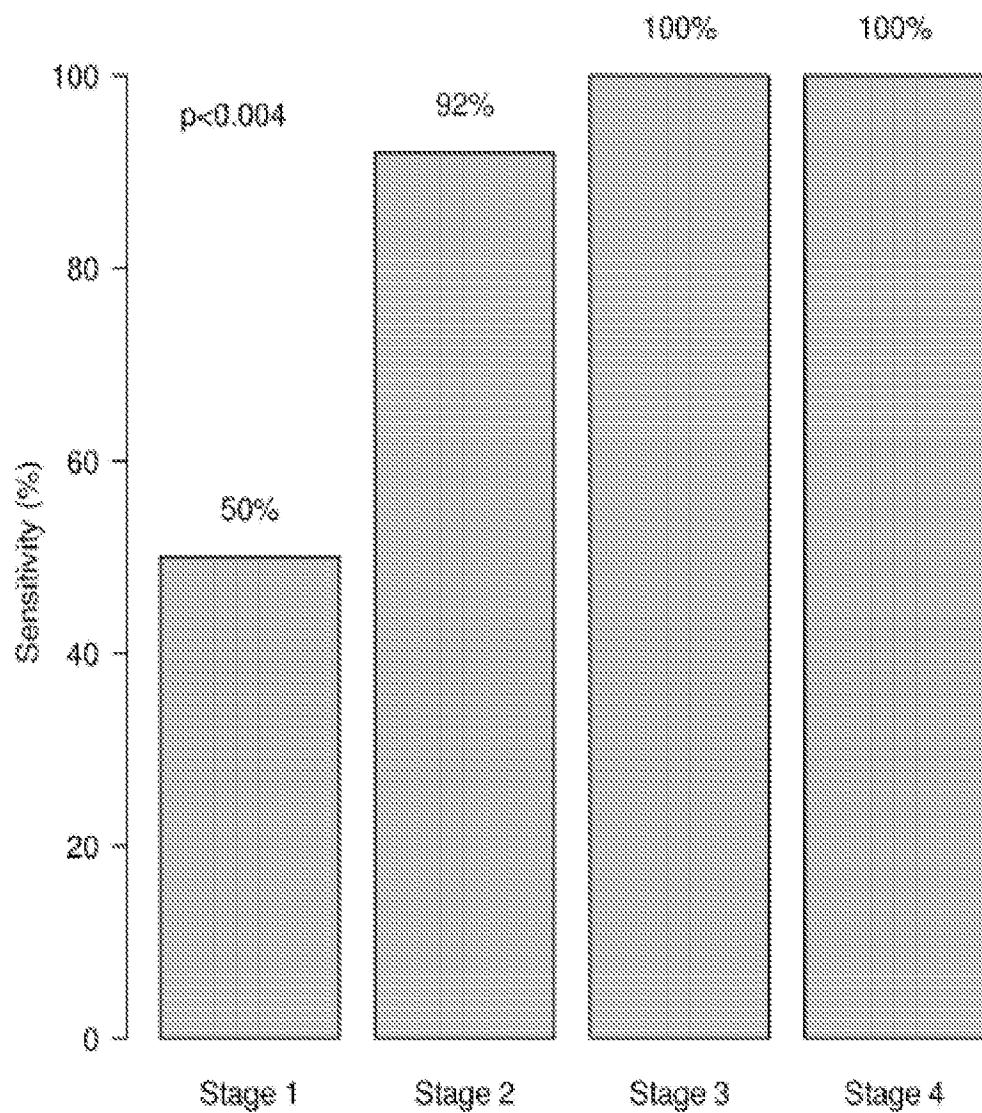
**FIG. 1**

FIG. 2

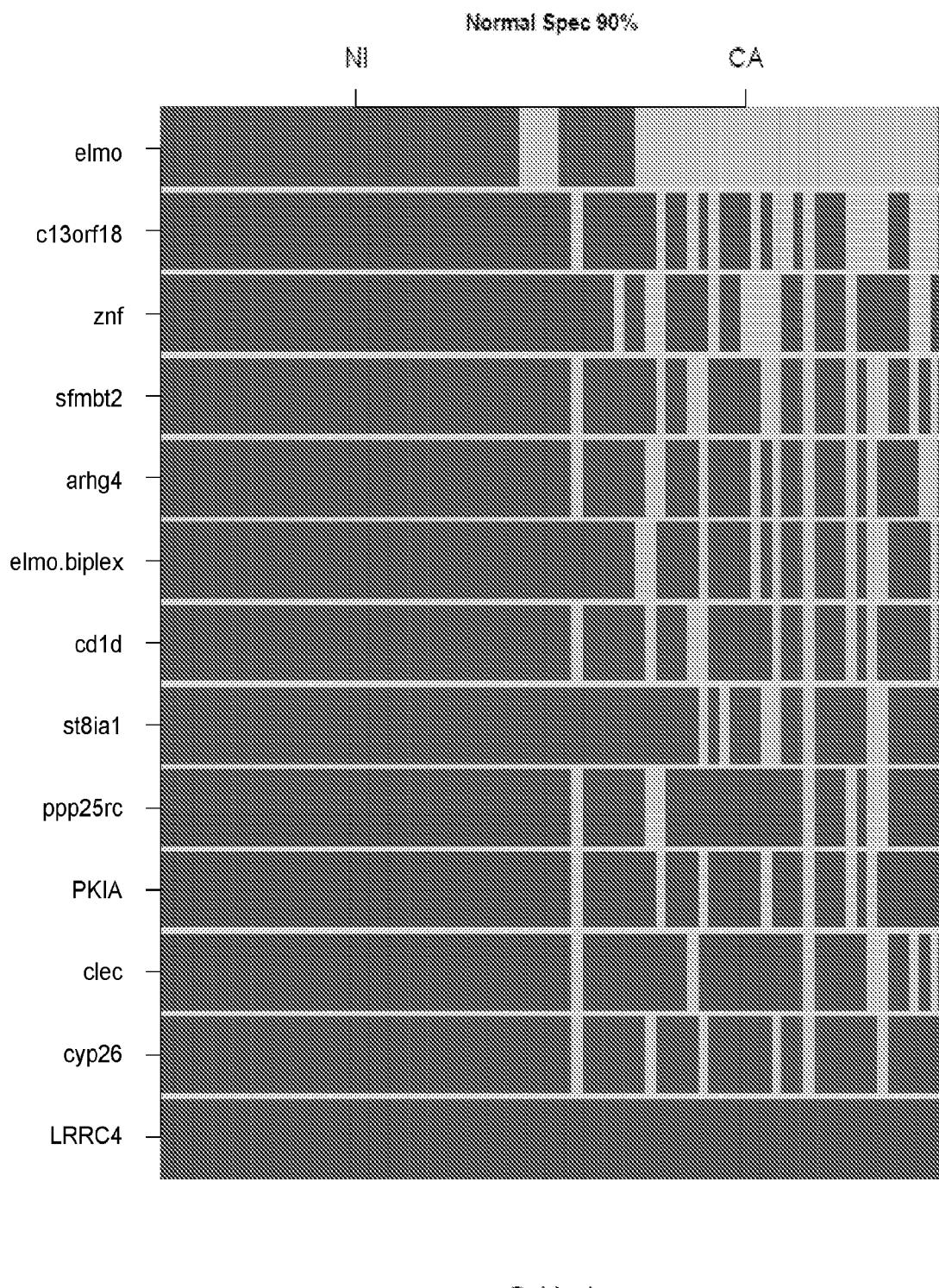
5'-G-CGCTG-TCT-BAG-2'-ACCGCTTTCGCTTACACTTGCTC-C6-3' (SEQ ID NO: 110)	4 5'-CGCCGAG (SEQ ID NO: 116)
5' 3' 00255 5' -GCTGAGCTT 3' -CG-CCTGCTGCTGCTT Arm 3	3 5'-GAGCGAG (SEQ ID NO: 117)
5' d-THX-TCT-BAG-1-ACCGCTTTCGCTTACACTTGCTC-C6-3' (SEQ ID NO: 112)	6 5'-GCAAGAC (SEQ ID NO: 118)
5' 3' 00256 5' -TAA TCTGAGCTT 3' -CG-CGCGCTGCGCTGCTT Arm 5	7 5'-GCCGTC (SEQ ID NO: 119)
5' d-THX-TCT-BAG-1-ACCGCTTTCGCTTACACTTGCTC-C6-3' (SEQ ID NO: 114)	for the methylation marks, A is HX, G is TAA, A is QUAGA,
5' 3' 00336 5' -HX TCTGAGCTT 3' -CG-CGCGCTGCGCTGCTT Arm 1	

**FIG. 3**

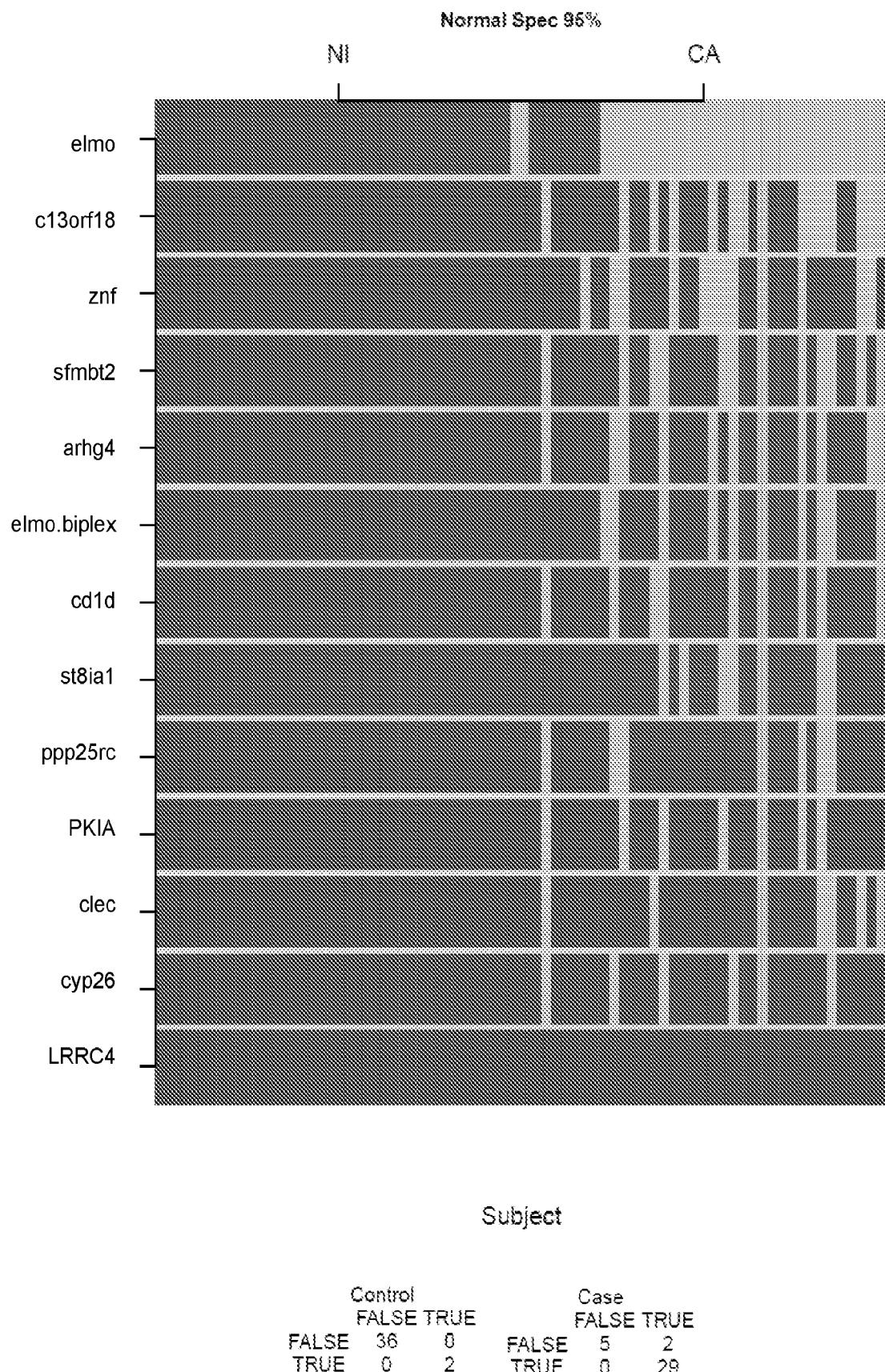
**FIG. 4A**

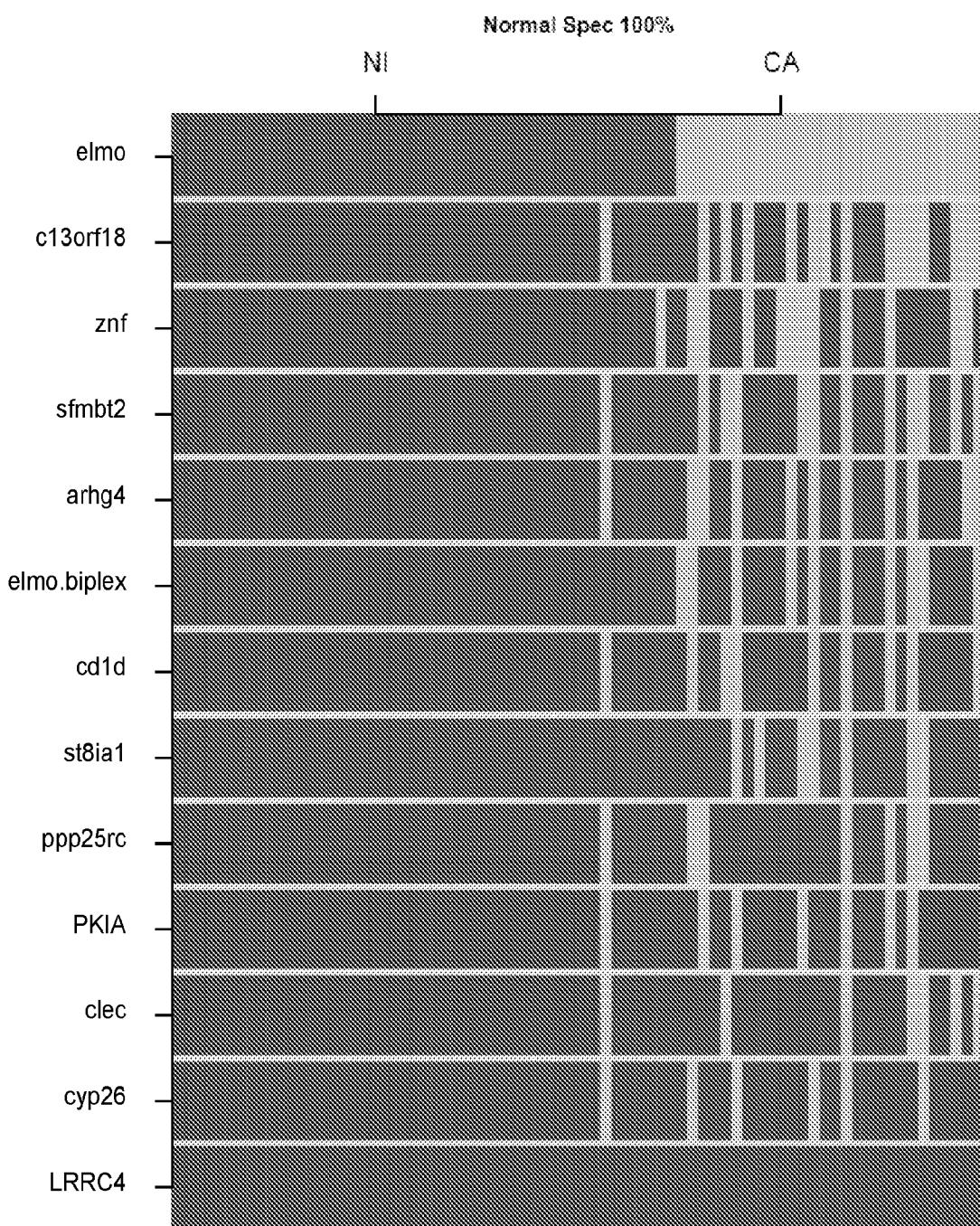
**FIG. 4B**

**FIG. 5A**



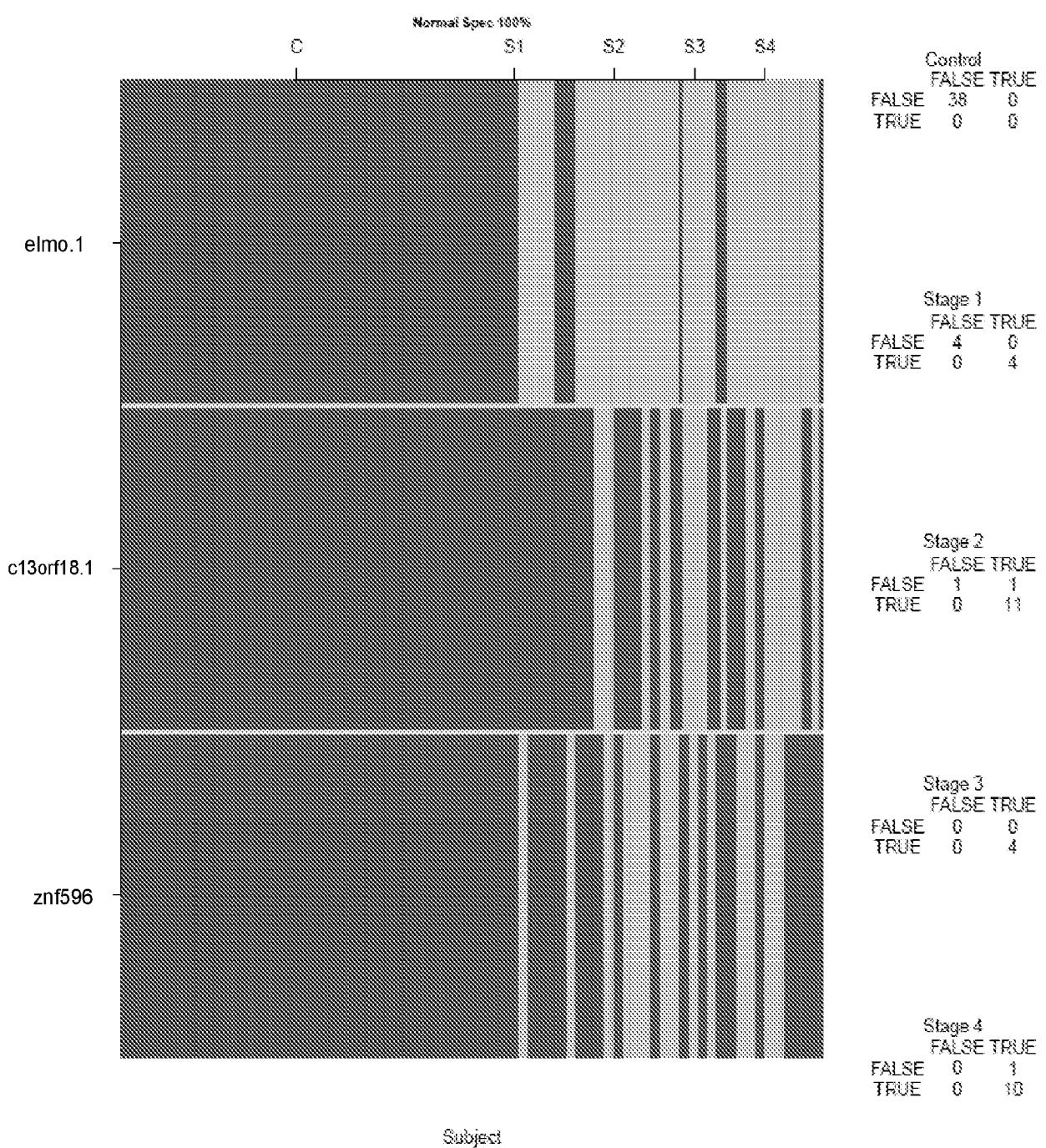
Control		Case		
	FALSE	TRUE	FALSE	
FALSE	34	0	5	2
TRUE	0	4	0	29

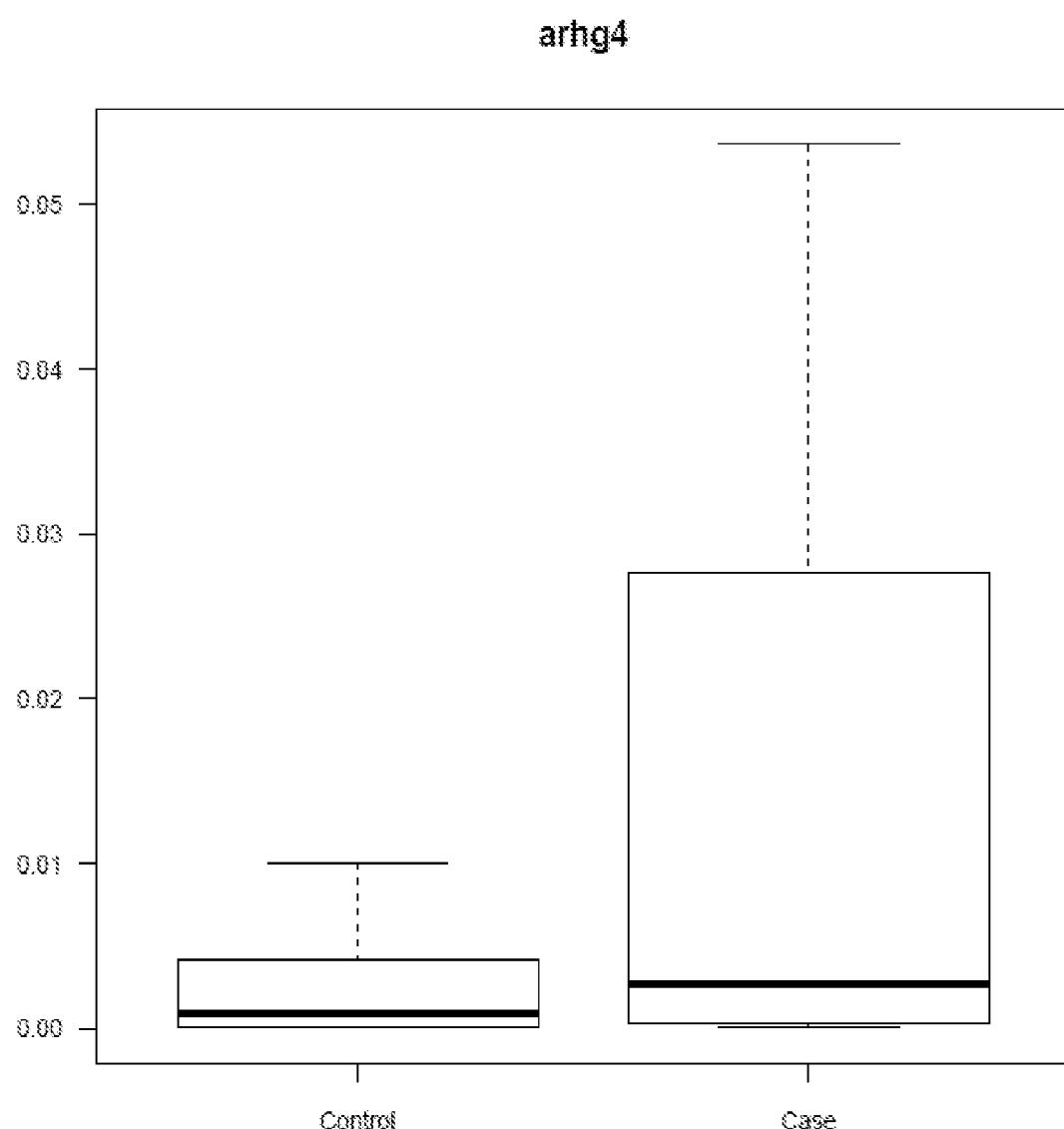
**FIG. 5B**

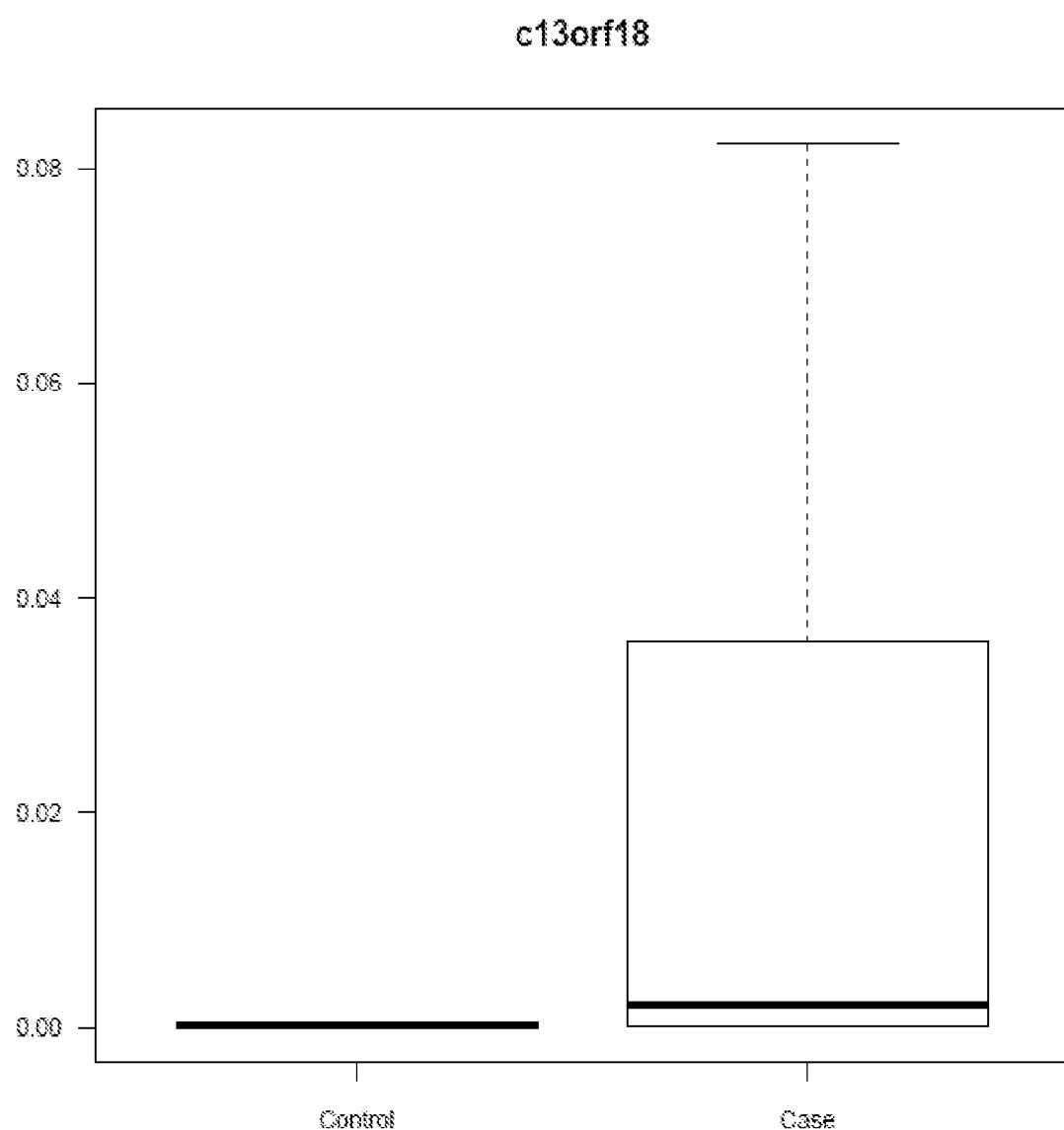
**FIG. 5C**

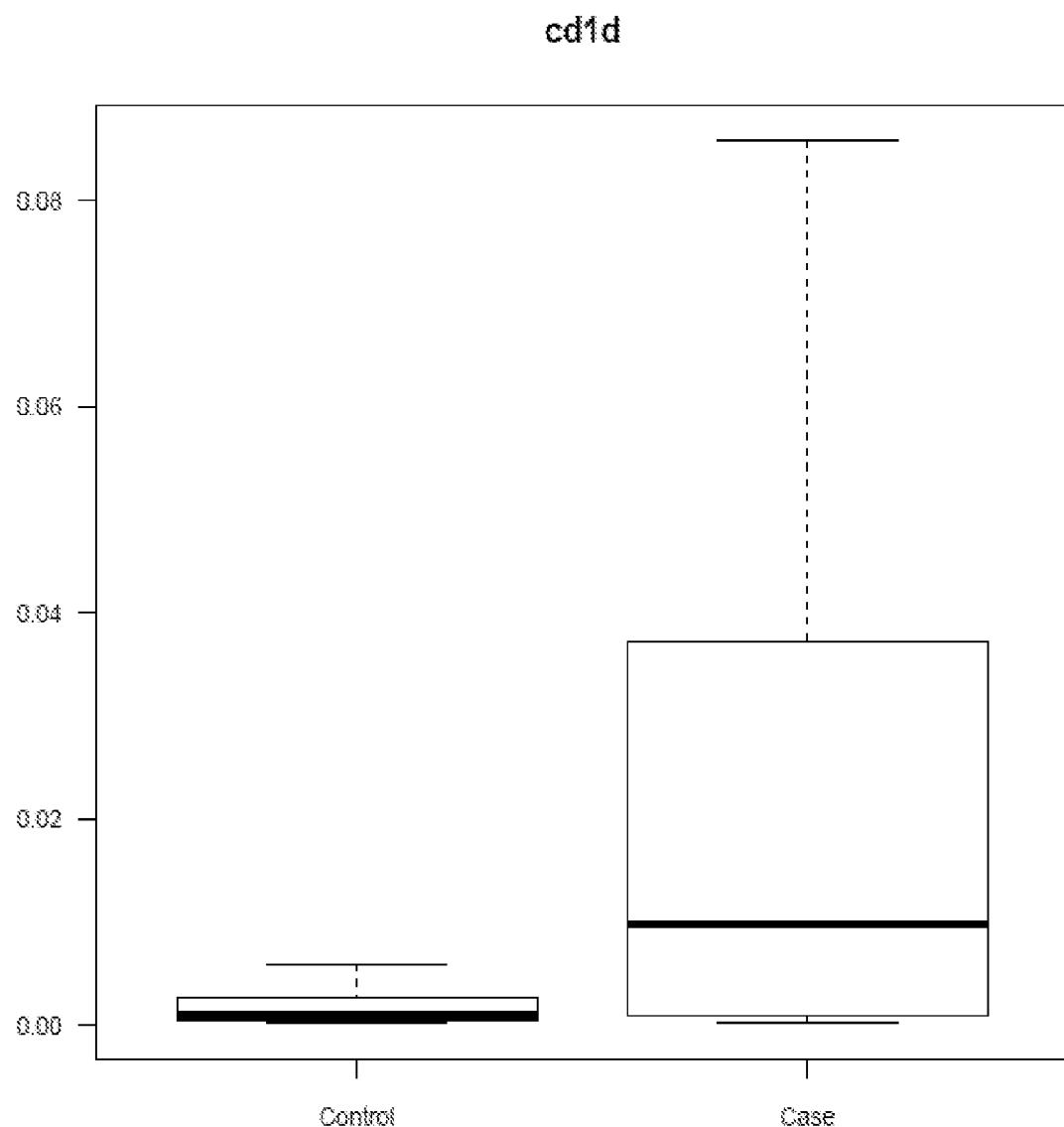
Subject

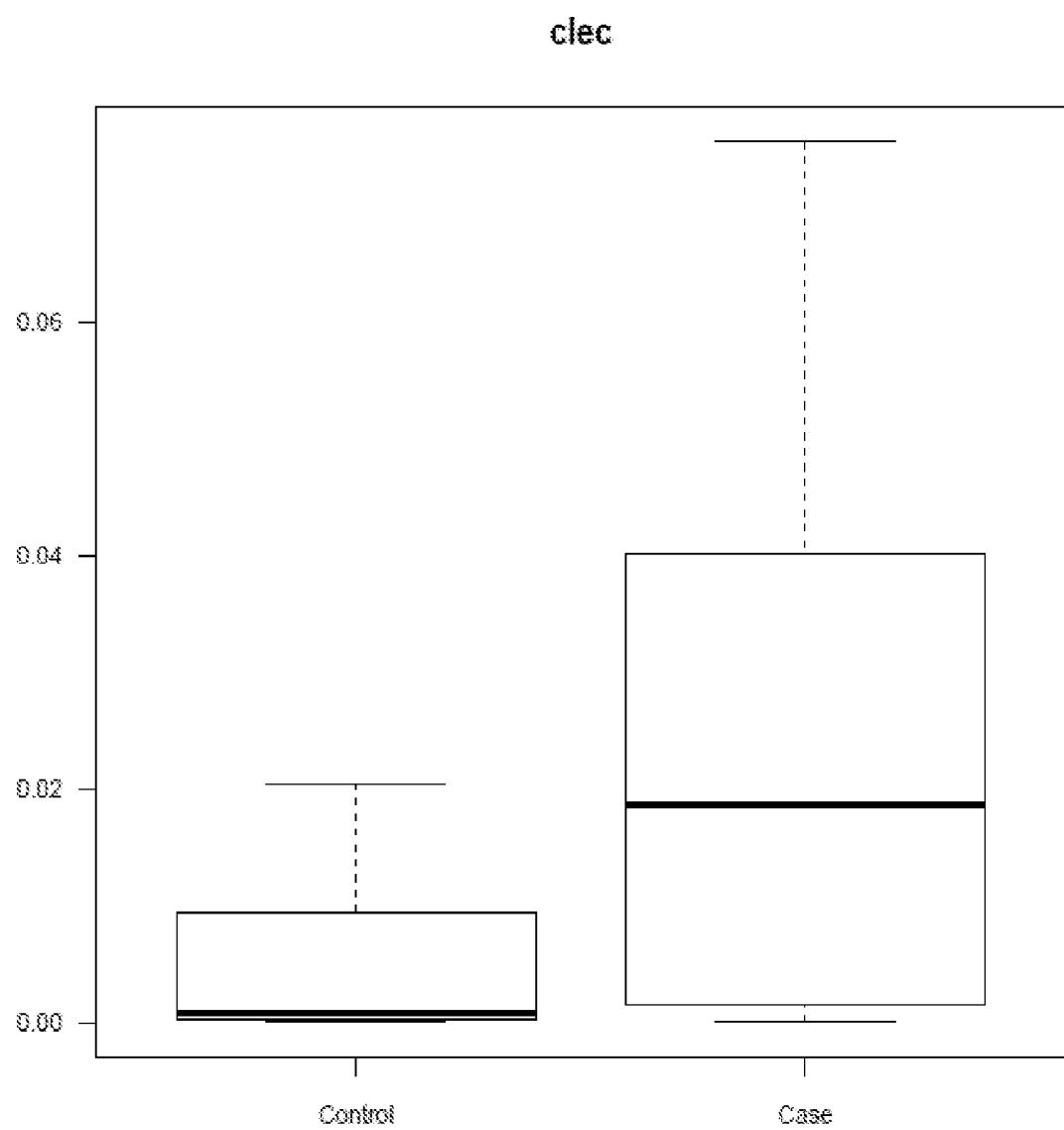
		Control		Case	
		FALSE	TRUE	FALSE	TRUE
FALSE	38	0		6	2
TRUE	0	0		0	28

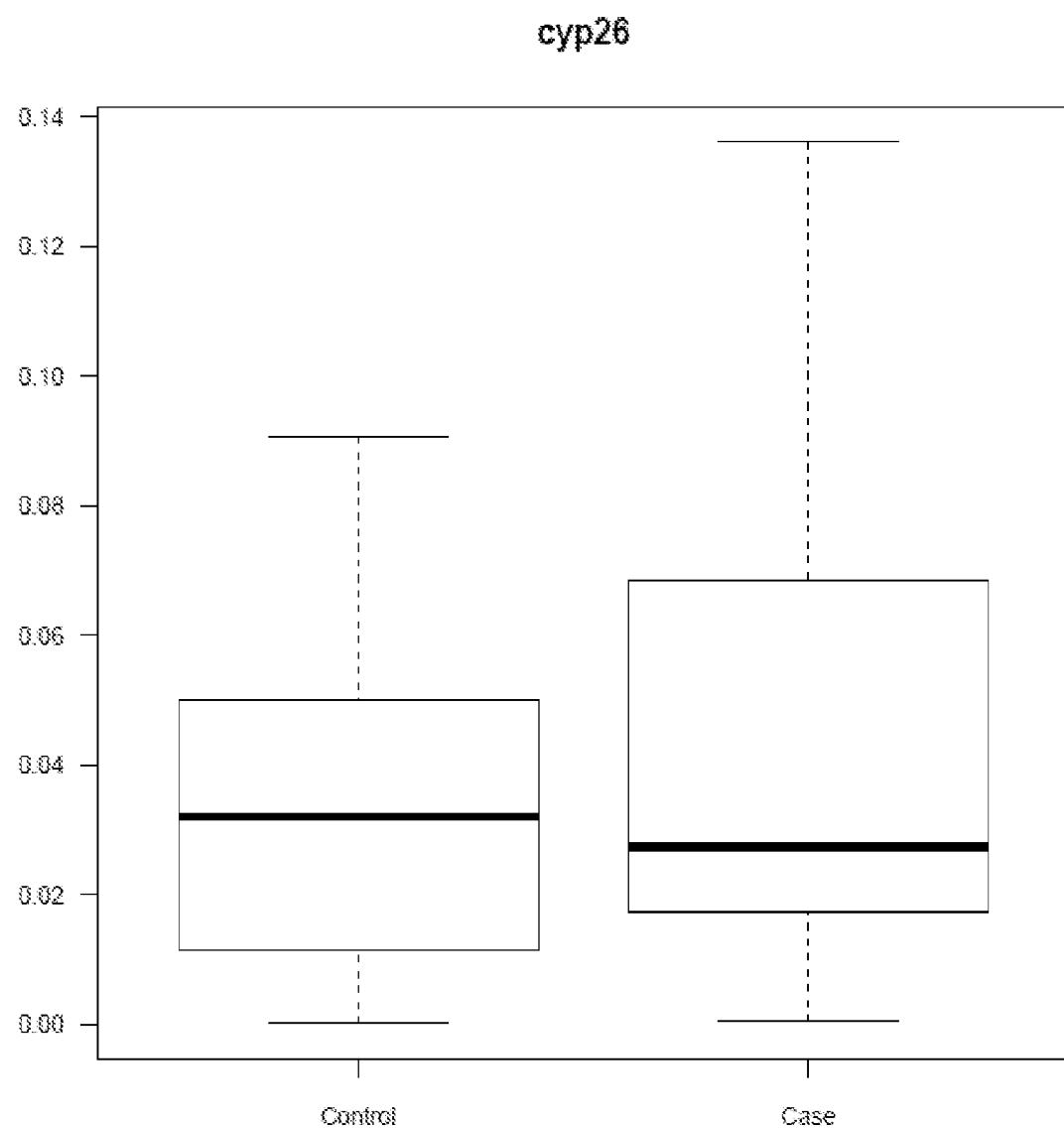
**FIG. 6**

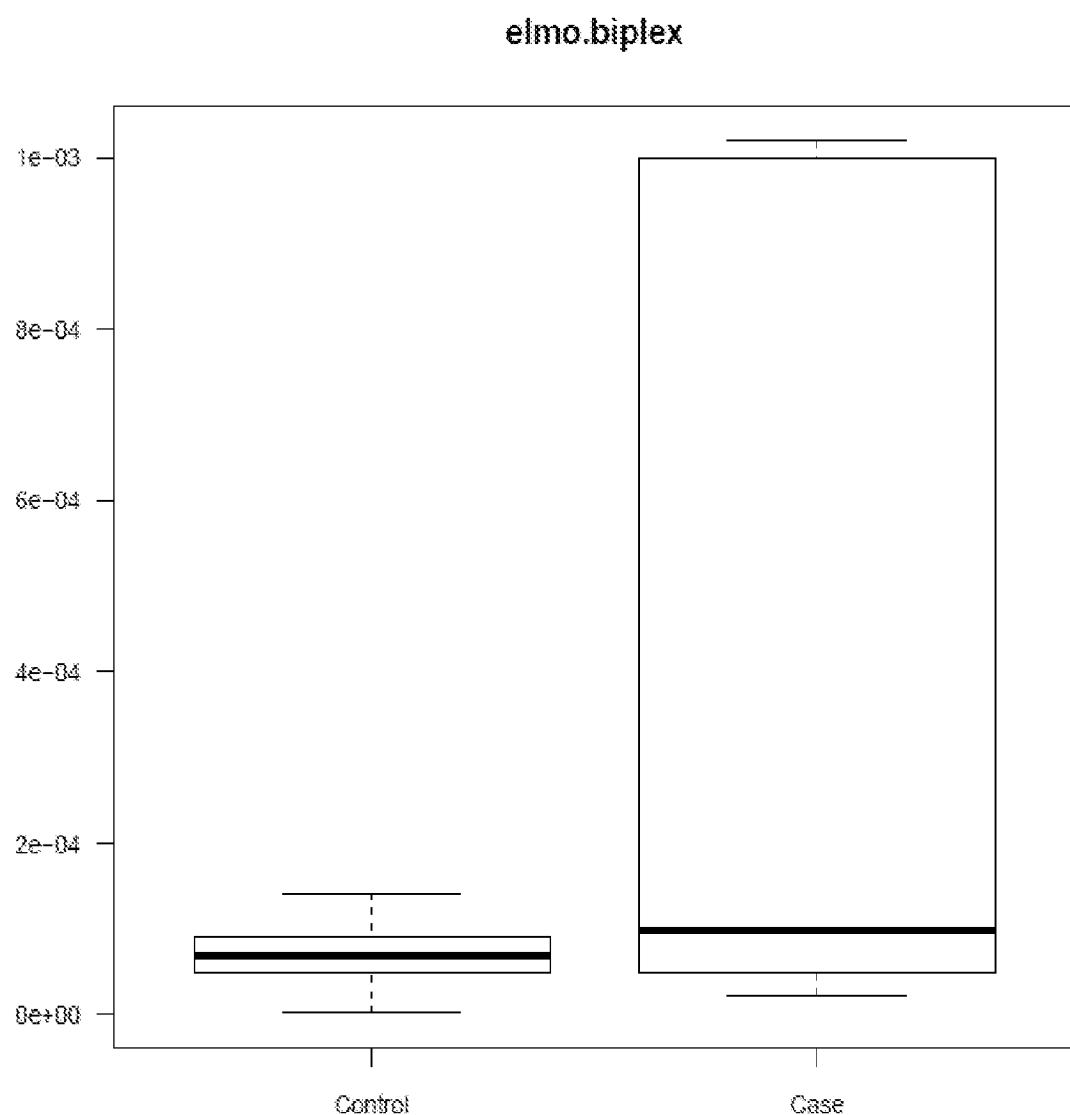
**FIG. 7A**

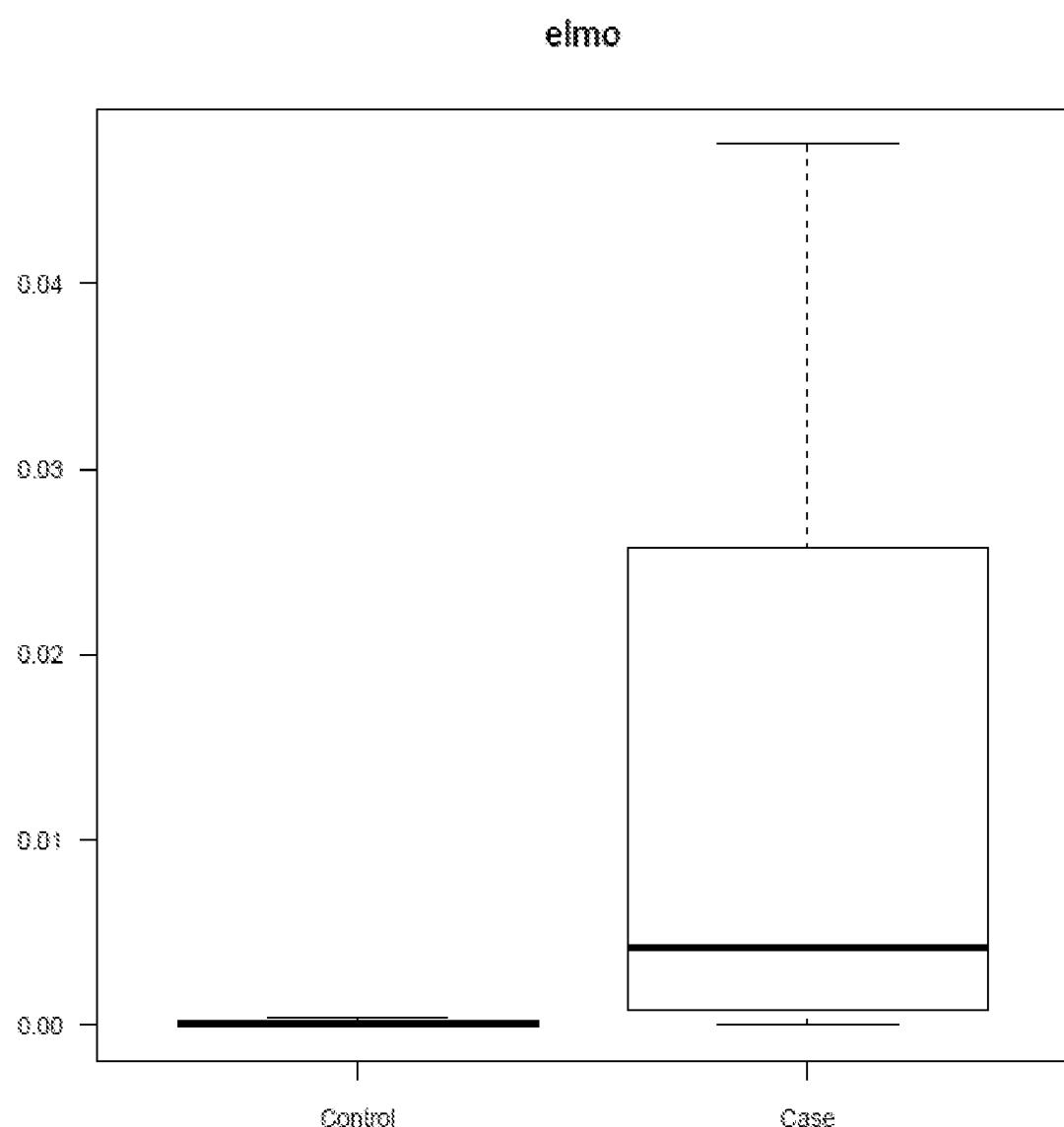
**FIG. 7B**

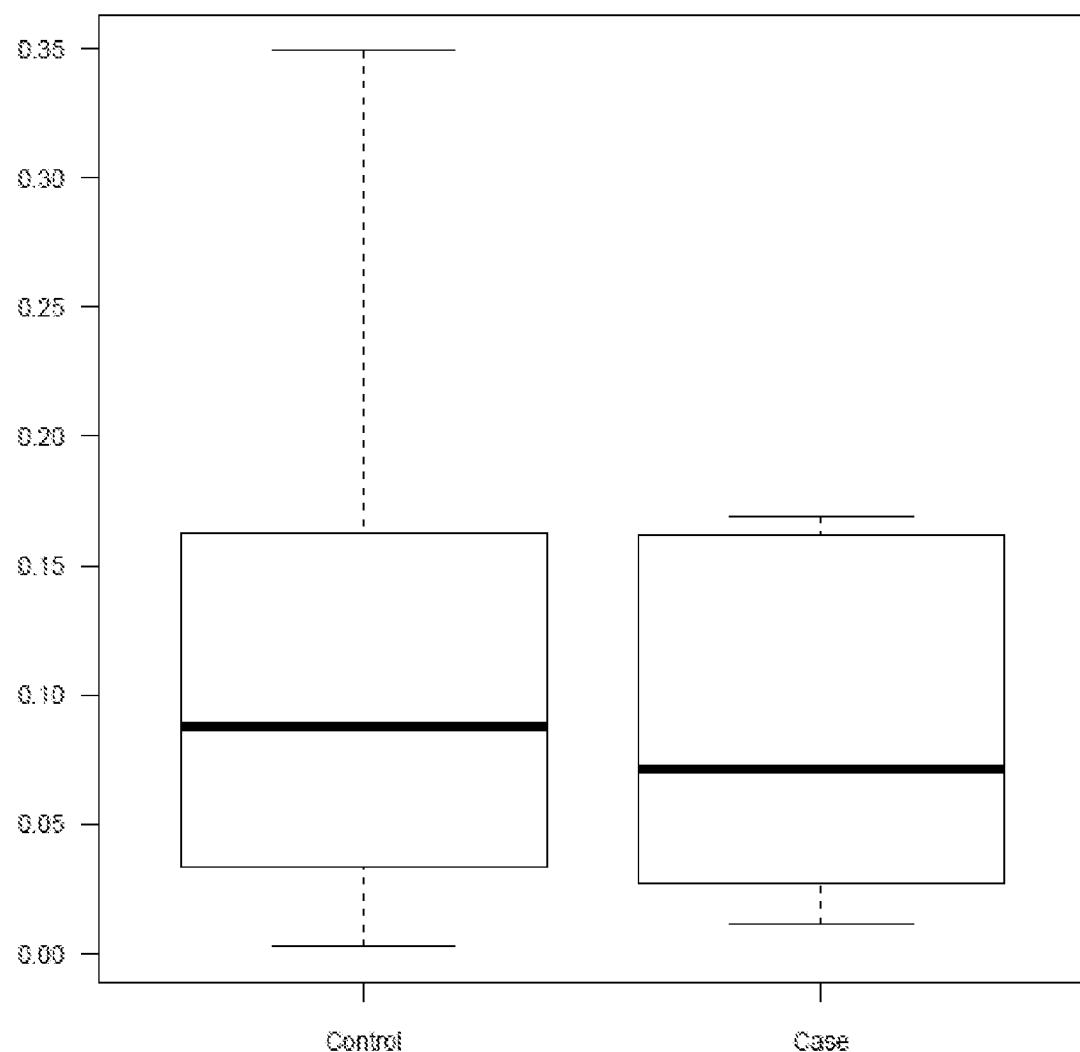
**FIG. 7C**

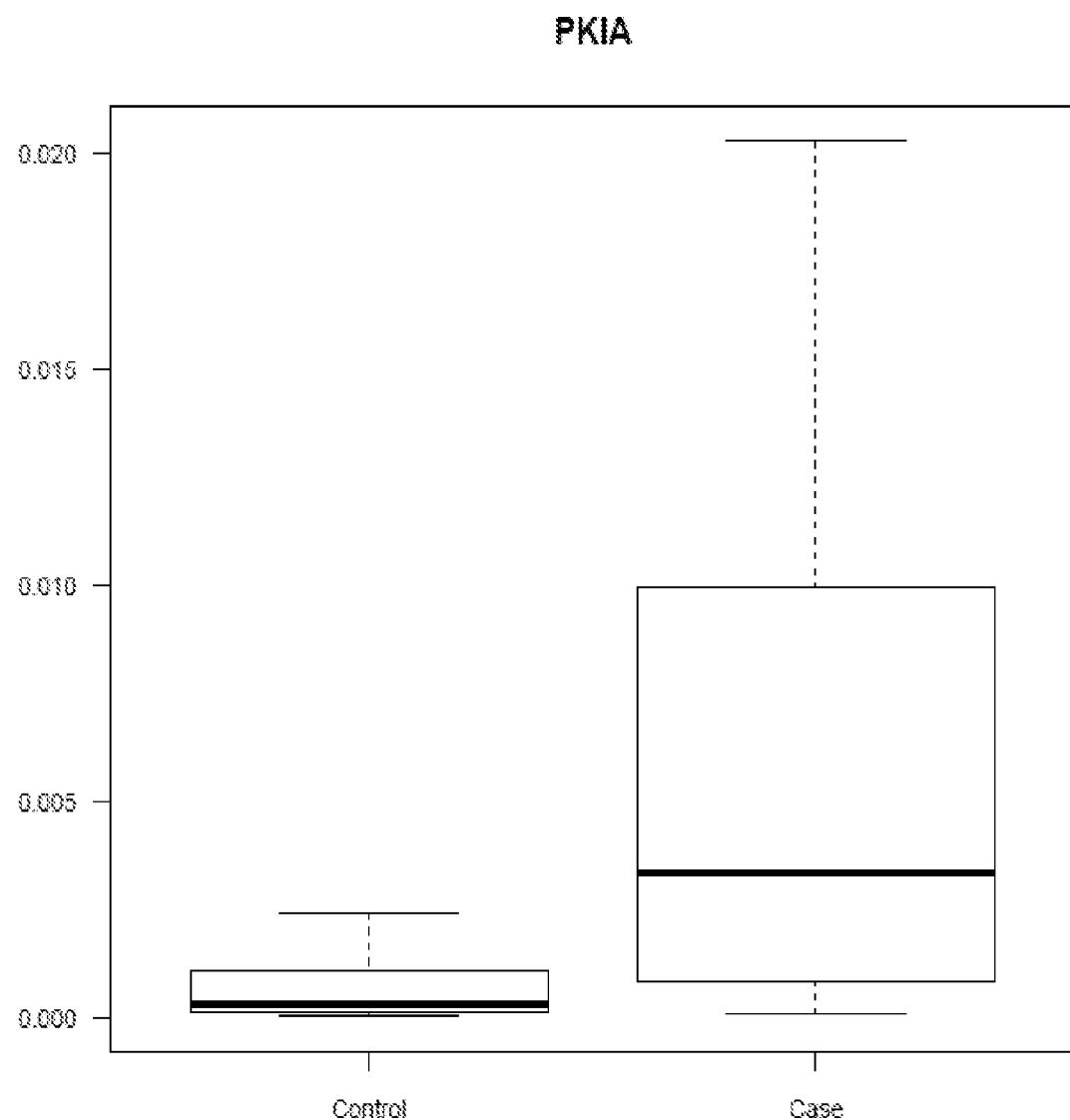
**FIG. 7D**

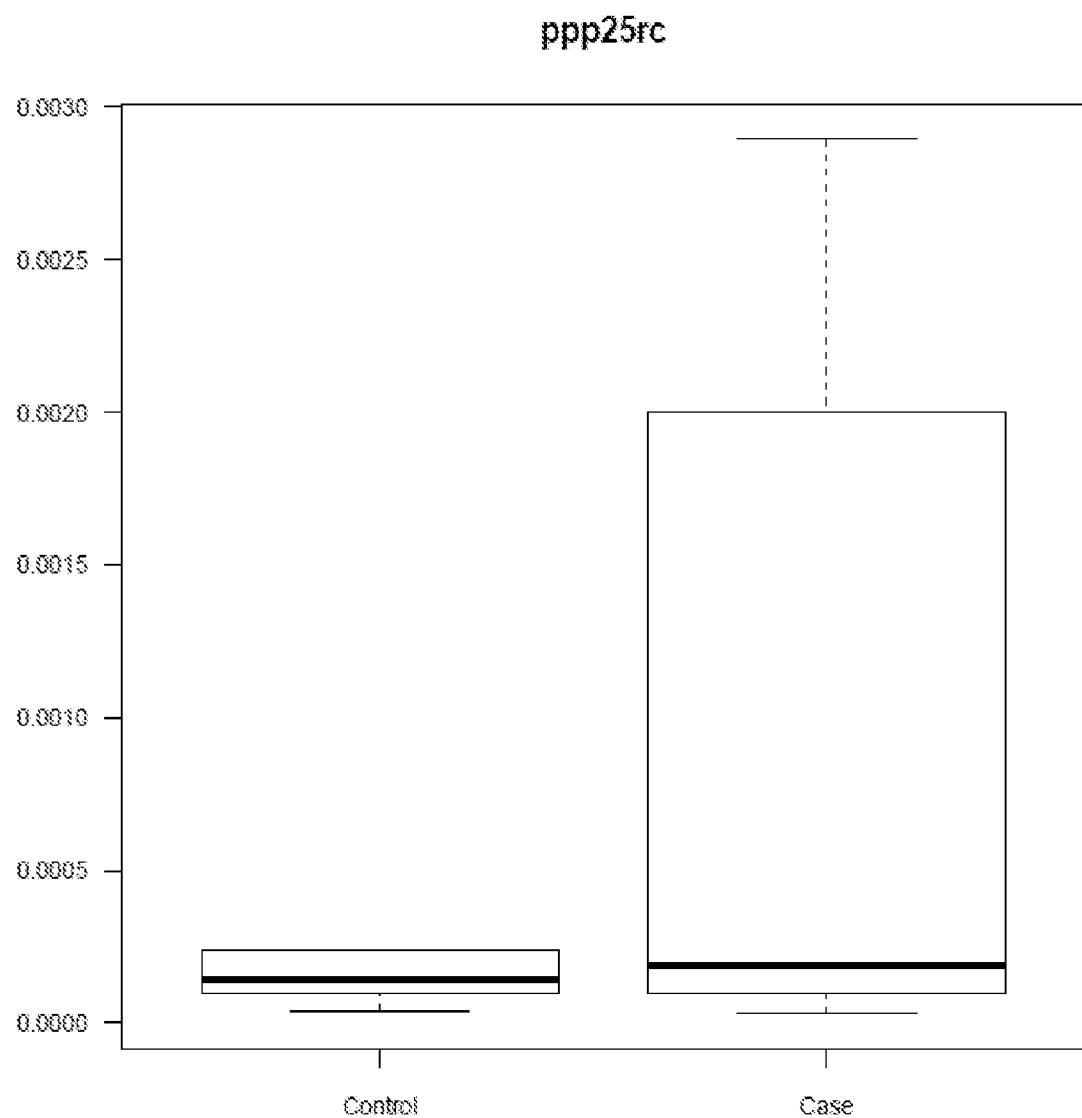
**FIG. 7E**

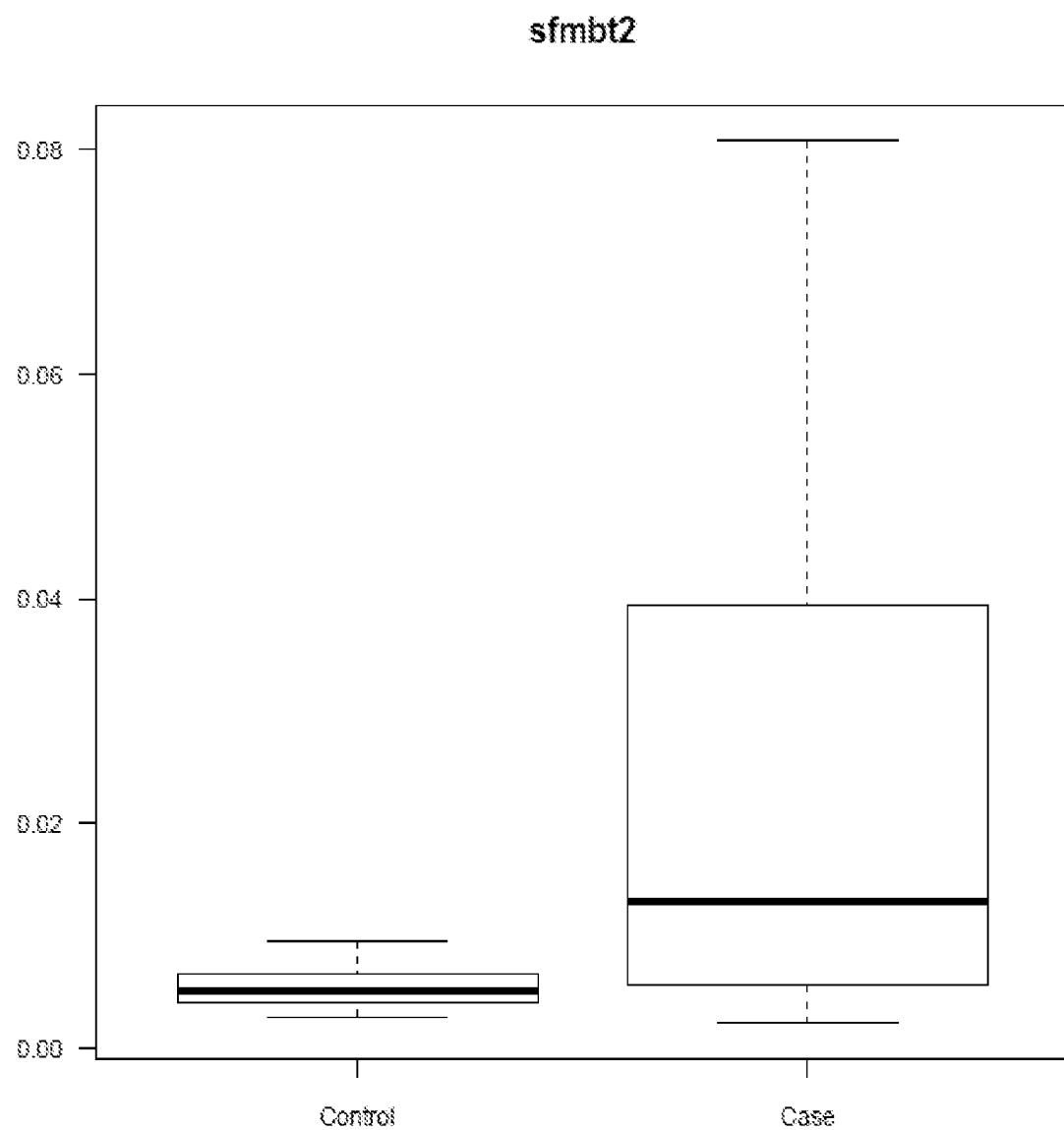
**FIG. 7F**

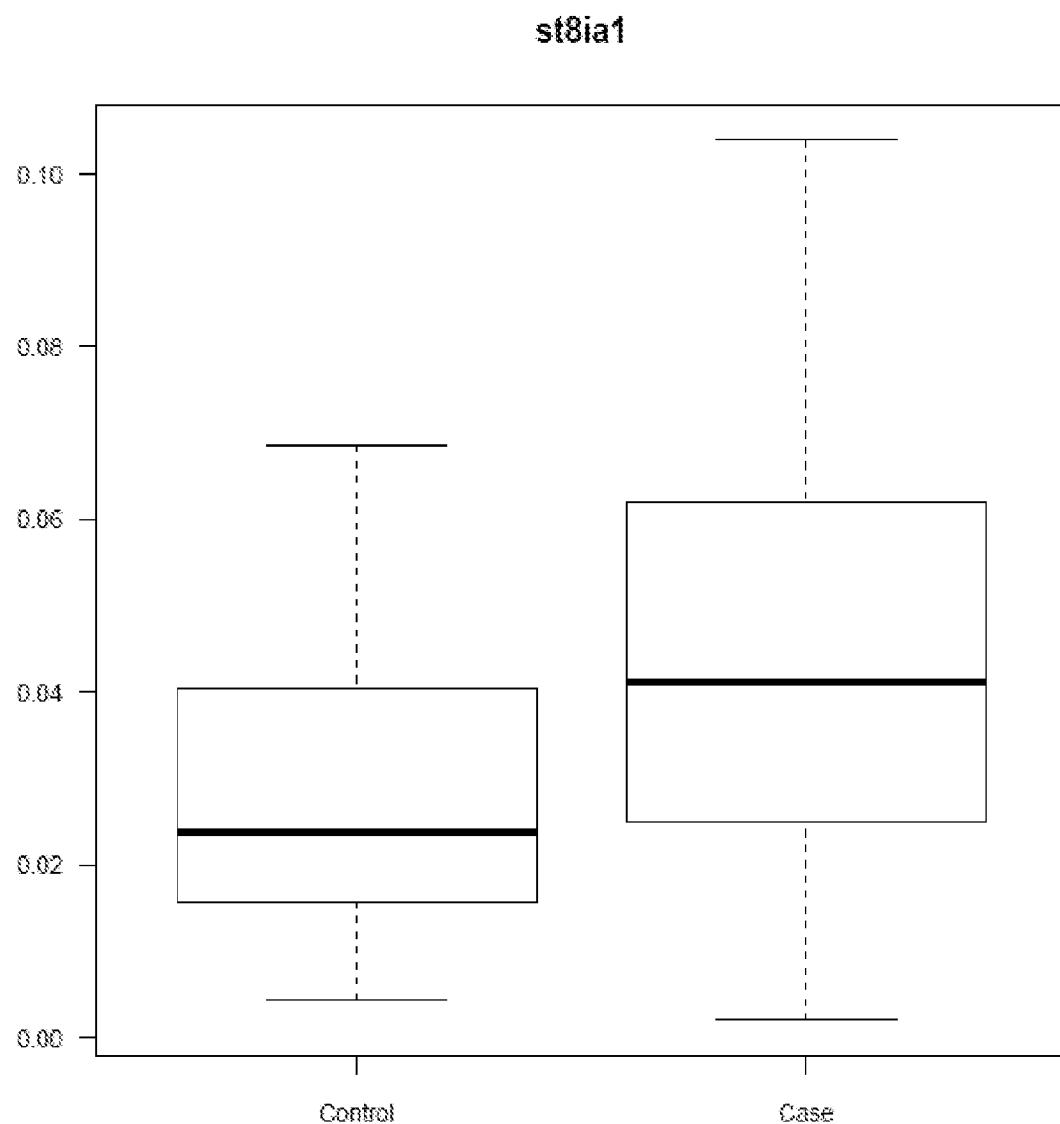
**FIG. 7G**

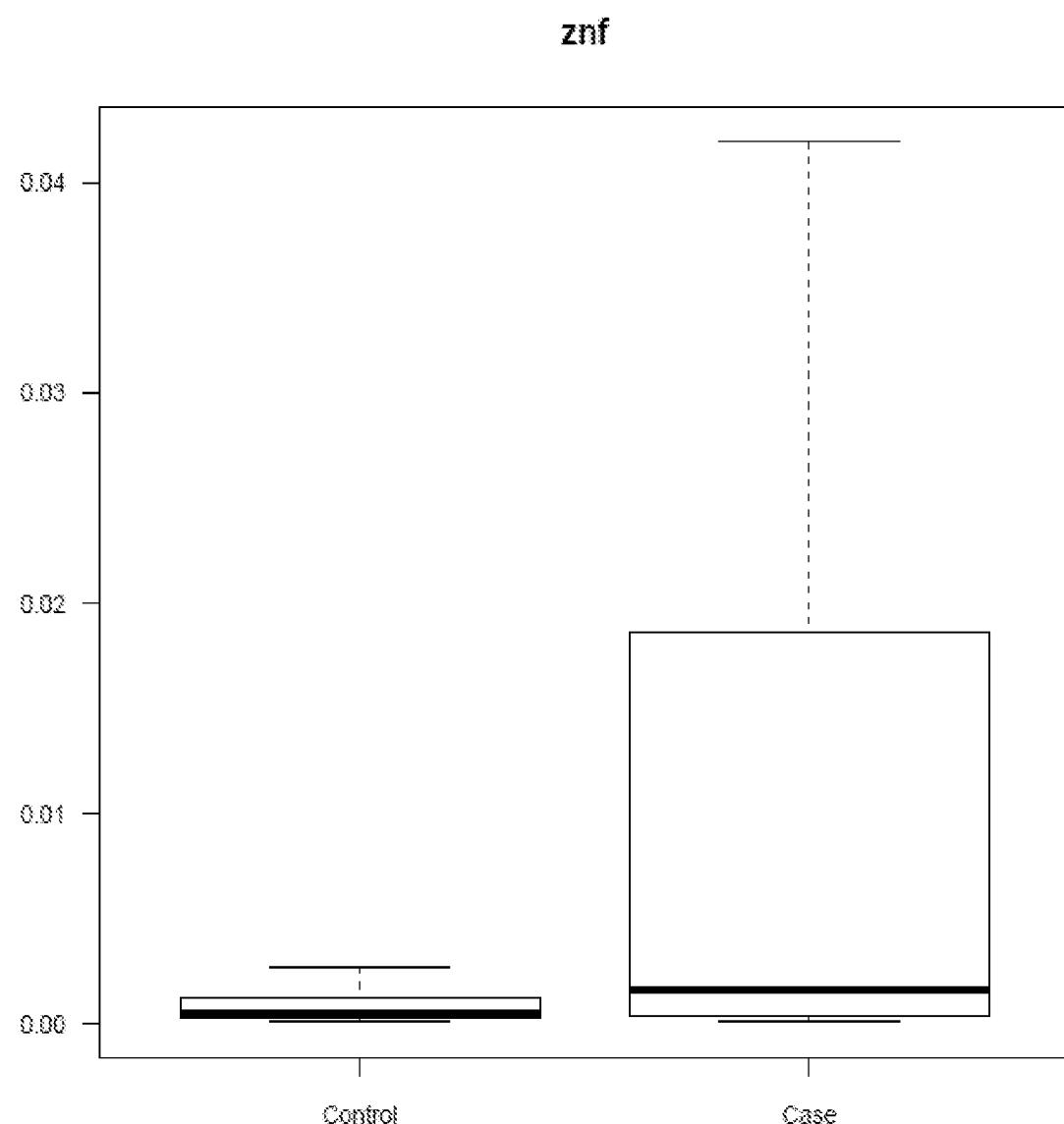
**FIG. 7H****LRRC4**

**FIG. 7I**

**FIG. 7J**

**FIG. 7K**

**FIG. 7L**

**FIG. 7M**

34505-WO-1-ORD\_ST25  
SEQUENCE LISTING

<110> MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH  
EXACT SCIENCES CORPORATION

<120> DETECTING GASTRIC NEOPLASM

<130> EXCTM-34505/WO-1/ORD

<150> US 62/212, 221

<151> 2015-08-31

<160> 119

<170> PatentIn version 3.5

<210> 1

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic

<400> 1

agtttggcgt agtcggtaga tc

22

<210> 2

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic

<400> 2

gcgcgcaaat accgaataaa cg

22

<210> 3

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic

<400> 3

gcgcgttagcg gcgttgc

17

<210> 4

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic

<400> 4

cccatatcgc ccgacgtaa

19

<210> 5  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 5  
 aaaacgtttt taggtatttg gtcgt

25

<210> 6  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 6  
 ataaaaaaaaac ctcctccgaa cccgc

25

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

<220>  
 <223> synthetic

<400> 7  
 ggagggagtt taacgggtta ggtac

25

<210> 8  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 8  
 cgtaaacaaa aaacaacgac gacg

24

<210> 9  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 9  
 gattttgttc gtcgttagtg c

21

<210> 10  
 <211> 19

<212> DNA  
 <213> Artifical Sequence

<220>  
 <223> synthetic

<400> 10  
 tctctaaacc cgcgAACG

19

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

<220>  
 <223> synthetic

<400> 11  
 tgtggaatcg gggtttgtgt tcgc

24

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

<220>  
 <223> synthetic

<400> 12  
 cccacccaac acaaaaaatc cgacg

25

<210> 13  
 <211> 23  
 <212> DNA  
 <213> Artifical Sequence

<220>  
 <223> synthetic

<400> 13  
 gacgtttgtc gtcgggttcg ttc

23

<210> 14  
 <211> 24  
 <212> DNA  
 <213> Artifical Sequence

<220>  
 <223> synthetic

<400> 14  
 aaaaaccctc cgctaccact tcgc

24

<210> 15  
 <211> 25  
 <212> DNA  
 <213> Artifical Sequence

<220>  
<223> synthetic

<400> 15  
ggtttttgg ttatccgga atcgt

25

<210> 16  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 16  
tataaaaacg cgcgtaatca acgct

25

<210> 17  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 17  
tgtttcgct gtcgttatat attacgtcgt

30

<210> 18  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 18  
gaactatccc cgaactccga ctcga

25

<210> 19  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 19  
tttaggaaag taaagcgtcg ttttc

25

<210> 20  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 20	aacgacgtct cgatacctac ga	22
<210> 21		
<211> 34		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 21	ttatatttt cgtttttagt aatttcgcgt tagc	34
<210> 22		
<211> 23		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 22	gaaaacccgc cgaaacattt cga	23
<210> 23		
<211> 30		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 23	tgttttcgcg gtcgttatat attacgtcgt	30
<210> 24		
<211> 25		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 24	gaactatccc cgaactccga ctcga	25
<210> 25		
<211> 31		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 25	cgggttttag cgatgttat ttttagttcg t	31

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

<220>  
 <223> synthetic

<400> 26  
 cctttcgtt cgtataaaat ttcgtt

26

<210> 27  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 27  
 cggggtaaa aaggtagcg atagtgatac

30

<210> 28  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 28  
 cgaaaaattc gaaacgaaac gtc

23

<210> 29  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 29  
 agtttggcgt agtcggtaga tc

22

<210> 30  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 30  
 gcgcgcaaat accgaataaa cg

22

<210> 31  
 <211> 23

<212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 31  
 gtttaatttt cggttcgtc gtc

23

<210> 32  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 32  
 cgctacgaaa cactccga

18

<210> 33  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 33  
 ttttgttggg tgggtatttt cgcg

24

<210> 34  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 34  
 acaaccctca aaaccctcg atatc

25

<210> 35  
 <211> 27  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 35  
 tttattttatg gattacgtt ttagcga

27

<210> 36  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
<223> synthetic

<400> 36  
gacgaccgaa cgatcacgca

20

<210> 37  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 37  
gcgtcgtag tagtacgaag c

21

<210> 38  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 38  
gcacctcaac gaaaacacccg a

21

<210> 39  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 39  
gcgcgtacgc gcgttcc

17

<210> 40  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 40  
cccatatcgc ccgacgtaa

19

<210> 41  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 41  
gtcgaaaaatgg ggcgttttag aggc

24

<210> 42  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 42  
taatcccaaa cgaaccgccc

20

<210> 43  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 43  
ggttttttgg ttatccgga atcgt

25

<210> 44  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 44  
tataaaaaacg cgcgtaatca acgct

25

<210> 45  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 45  
ttgagatgtt gggtgaaggc gattc

25

<210> 46  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 46  
cgctaacgcg aaaaaataat tcgacg

26

34505-WO-1-0RD\_ST25

<210> 47  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 47  
cggttttatt tattatgatt cgttagcgg

28

<210> 48  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 48  
cgactaccct aaacaacgca tcgc

24

<210> 49  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 49  
cgactaccct aaacaacgca tcgc

24

<210> 50  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 50  
cccgccgaat actcgatcaa ctcg

24

<210> 51  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 51  
ggtcgcgtcg ttttcgatc

19

<210> 52  
<211> 17

<212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 52  
 gccgcaaacg ccgacga

17

<210> 53  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 53  
 tgtatatttc gaggcggtt cg

23

<210> 54  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 54  
 cgcccccaaa ataaaaaaaaac gaa

23

<210> 55  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 55  
 ttttcgttc gtttttttat ttttcga

28

<210> 56  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 56  
 aacgcataacc atcgacttca actacgac

28

<210> 57  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
<223> synthetic

<400> 57  
cggtttcgt tcgttttc g

21

<210> 58  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 58  
ccgccttcta cgcgacta

18

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

<220>  
<223> synthetic

<400> 59  
gtattttaa tatttggta gtgcg

25

<210> 60  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 60  
tactaaaacg tacgataaac atcgt

25

<210> 61  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 61  
tcgattttat ttttgggtc gtttagatt cgc

33

<210> 62  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 62	gaaaaaaacta aaaaacgaca aaaaaacccg acg	33
<210> 63		
<211> 25		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 63	ggagggagtt taacgggtta ggtac	25
<210> 64		
<211> 24		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 64	cgtaaacaaa aaacaacgac gacg	24
<210> 65		
<211> 25		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 65	attgcgtcgt ttttaggta acgtt	25
<210> 66		
<211> 23		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 66	cggaaaatacc tatctcctcg act	23
<210> 67		
<211> 28		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> synthetic		
<400> 67	ggaaacgagtg atagtcggat agttcgtc	28

<210> 68  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 68  
 cgcccgaaaa cgaccccg

18

<210> 69  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 69  
 gggagtgagg gtagggcgtt c

21

<210> 70  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 70  
 ctcgcaaccc cttcgaattt acccg

25

<210> 71  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 71  
 gcgagttgat cgagtattcg g

21

<210> 72  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 72  
 ctaatcaacg caaatccccc g

21

<210> 73  
 <211> 22

<212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 73  
 ccacggacgc gcacgccc ac

22

<210> 74  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 74  
 cgttagtttt tagaagtttt ttaggaaagt

30

<210> 75  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 75  
 cgacgtctcg atacctacga

20

<210> 76  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 76  
 ccacggacga cgaaaacgac gc

22

<210> 77  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 77  
 gcgggagttt ggcgtag

17

<210> 78  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
<223> synthetic

<400> 78  
cgcgcaaata ccgaataaac g

21

<210> 79  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 79  
ccacggacgg tcggtagatc gc

22

<210> 80  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 80  
gttttcgggt ttcgggtttt g

21

<210> 81  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 81  
cgctaaaaat acgacgcccc

20

<210> 82  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 82  
ccacggacgg cgttttattc gc

22

<210> 83  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 83	19
gcgtttaggt tggtcggag	
<210> 84	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 84	19
acgcactcaa cctacgaac	
<210> 85	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 85	22
ccacggacgc cgaaaaacta cc	
<210> 86	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 86	22
agagttcggc gtttagagtt ag	
<210> 87	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 87	21
tcgaaacctatccgcgaa c	
<210> 88	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 88	22
ccacggacgg cggttttcg ac	

34505-W0-1-ORD\_ST25

<210> 89	
<211> 27	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 89	
tttgggggg tgattaggtg tttaaga	27
<210> 90	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 90	
caccaacctc ataaccttat c	21
<210> 91	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 91	
gacgcggaga tagtgttgt gc	22
<210> 92	
<211> 29	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 92	
ttcgatttta tttttgttgt cgttgtaga	29
<210> 93	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> synthetic	
<400> 93	
acgacaaaaa aacccgacga	20
<210> 94	
<211> 21	

<212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 94  
 cgccgaggat tcgcgtgggt c

21

<210> 95  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 95  
 cgttcgcgtt atttatttcg gcg

23

<210> 96  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 96  
 gctcctaatt ctcataacg tcgt

24

<210> 97  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 97  
 cgccgagggc ggcgtttgc c

21

<210> 98  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 98  
 gaagagtgcg taggttagag g

21

<210> 99  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence

<220>  
<223> synthetic

<400> 99  
catatcgccc gacgtaaaaa cc

22

<210> 100  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 100  
cgccgaggct cgcgaaacgc c

21

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

<220>  
<223> synthetic

<400> 101  
aaatatgtgg ttcgtggcgt t

21

<210> 102  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 102  
acgcaacaac gcgaaaaac

19

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

<220>  
<223> synthetic

<400> 103  
cgccgaggcg acgaaaaacg c

21

<210> 104  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 104  
cgttcgcgtt atttatttcg gcg

23

<210> 105  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 105  
gctcctaatt ctcataaacc tcgt

24

<210> 106  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 106  
cgccgagggc ggcgttttgc c

21

<210> 107  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 107  
gcgttaattt cgcgaggta

19

<210> 108  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 108  
acaatactct tatataattaa cgccgctc

28

<210> 109  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 109  
cgccgaggag gcgacggagg c

21

34505-WO-1-0RD\_ST25

<210> 110  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 110  
agccggtttt ccggctgaga ctccgcgtc

29

<210> 111  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 111  
tctagccgtt tttccggctg agactccgcg tc

32

<210> 112  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 112  
agccggtttt ccggctgaga cgtccgtgg

29

<210> 113  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 113  
tctagccgtt tttccggctg agacgtccgt gg

32

<210> 114  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> synthetic

<400> 114  
agccggtttt ccggctgaga cctcggcg

28

<210> 115  
<211> 31

<212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 115  
 tctagccggt tttccggctg agacctcgcc g

31

<210> 116  
 <211> 8  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 116  
 cgccgagg

8

<210> 117  
 <211> 9  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 117  
 gacgcggag

9

<210> 118  
 <211> 9  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 118  
 ccacggacg

9

<210> 119  
 <211> 8  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> synthetic

<400> 119  
 gcgcggtcc

8