(12) STANDARD PATENT (11) Application No. **AU 2007237444 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Methods and nucleic acids for the detection of colorectal cell proliferative disorders

(51) International Patent Classification(s)
C12Q 1/68 (2006.01)

(21) Application No: **2007237444** (22) Date of Filing: **2007.04.17**

(87) WIPO No: **WO07/118704**

(30) Priority Data

(31) Number	(32) Date	(33) Country
PCT/US2006/014131	**2006.04.17**	**US**
06090119.6	**2006.06.27**	**EP**

(43) Publication Date: **2007.10.25**
(44) Accepted Journal Date: **2013.05.23**

(71) Applicant(s)
Epigenomics AG

(72) Inventor(s)
Lofton-Day, Catherine;Ebert, Matthias

(74) Agent / Attorney
Pizzeys Patent and Trade Mark Attorneys, GPO Box 1374, BRISBANE, QLD, 4001

(56) Related Art
US 2005/0064401 A1
WO 2005/123945 A2
WO 2006/113466 A2
WO 2004/074441 A2

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 October 2007 (25.10.2007)

PCT

(10) International Publication Number
WO 2007/118704 A3

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/EP2007/003380

(22) International Filing Date: 17 April 2007 (17.04.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/US2006/014131 17 April 2006 (17.04.2006) US
06090119.6 27 June 2006 (27.06.2006) EP

(71) Applicant *(for all designated States except US)*: **EPIGENOMICS AG** [DE/DE]; Kleine Präsidentenstrasse 1, 10178 Berlin (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LOFTON-DAY, Catherine** [US/US]; 4816 Linden Ave N., Seattle, WA 98103 (US). **EBERT, Matthias** [DE/DE]; Adam-Berg-Strasse 176, 81735 München (DE).

(74) Agents: **KRAUSS, Jan, B.** et al.; Boehmert & Boehmert, Anwaltssozietät, Pettenkoferstrasse 20-22, 80336 Munich (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
24 April 2008

(54) Title: METHODS AND NUCLEIC ACIDS FOR THE DETECTION OF COLORECTAL CELL PROLIFERATIVE DISORDERS

(57) Abstract: The invention provides methods, nucleic acids and kits for detecting, or for distinguishing between or among colorectal cell proliferative disorders. The invention discloses genomic sequences the methylation patterns of which have utility for the improved detection of and differentiation between said class of disorders, thereby enabling the improved diagnosis and treatment of patients.

WO 2007/118704 A3

METHODS AND NUCLEIC ACIDS FOR THE DETECTION OF COLORECTAL CELL PROLIFERATIVE DISORDERS

FIELD OF THE INVENTION

The present invention relates to genomic DNA sequences that exhibit altered expression patterns in disease states relative to normal. Particular embodiments provide, *inter alia*, novel methods, nucleic acids, nucleic acid arrays and kits useful for detecting, or for detecting pre-cancerous colorectal lesions. Preferably, the methods, nucleic acids, nucleic acid arrays and kits for the screening of individuals to identify those at risk of developing colorectal carcinoma.

BACKGROUND

Incidence and diagnosis of cancer. Cancer is the second leading cause of death of the United States. Mortality rates could be significantly improved if current screening methods would be improved in terms of patient compliance, sensitivity and ease of screening. Current recommended methods for diagnosis of cancer are often expensive and are not suitable for application as population wide screening tests.

Hepatocellular cancer (HCC) is the fourth most common cancer in the world, its incidence varies from 2.1 per 100,000 in North America to 80 per 100,000 in China. In the United States, it is estimated that there will be 17,550 new cases diagnosed in 2005 and 15,420 deaths due to this disease. Ultrasound of the liver, alpha fetoprotein levels and conventional CT scan are regularly obtained in the diagnostic evaluation of HCC (hepatocellular cancer or primary liver cancer), but they are often too insensitive to detect multi-focal small lesions and for treatment planning.

In the United States the annual incidence of colorectal cancer is approximately 150,000, with 56,600 individuals dying form colorectal cancer each year. The lifetime risk of colorectal cancer in the general population is about 5 to 6 percent. Despite intensive efforts in recent years in screening and early detection of colon cancer, until today most cases are diagnosed in an advanced stage with regional or distant metastasis. While the therapeutic options include surgery and adjuvant or palliative chemotherapy, most patients die from progression of their cancer within a few months. Identifying the molecular changes that underlie the development of colon cancer may help to develop new monitoring, screening, diagnostic and therapeutic options that could improve the overall poor prognosis of these patients.

The current guidelines for colorectal screening according to the American Cancer Society utilizes one of five different options for screening in average risk individuals 50 years of age or older. These options include 1) fecal occult blood test (FOBT) annually, 2) flexible sigmoidoscopy every five

years, 3) annual FPBT plus flexible sigmoidoscopy every five years, 4) double contrast barium enema (DCBE) every five years or 5) colonoscopy every ten years. Even though these testing procedures are well accepted by the medical community, the implementation of widespread screening for colorectal cancer has not been realized. Patient compliance is a major factor for limited use due to the discomfort or inconvenience associated with the procedures. FOBT testing, although a non-invasive procedure, requires dietary and other restrictions 3-5 days prior to testing. Sensitivity levels for this test are also very low for colorectal adenocarcinoma with wide variability depending on the trial. Sensitivity measurements for detection of adenomas is even less since most adenomas do not bleed. In contrast, sensitivity for more invasive procedures such as sigmoidoscopy and colonoscopy are quite high because of direct visualization of the lumen of the colon. No randomized trials have evaluated the efficacy of these techniques, however, using data from case-control studies and data from the National Polyp Study (U.S.) it has been shown that removal of adenomatous polyps results in a 76-90% reduction in CRC incidence. Sigmoidoscopy has the limitation of only visualizing the left side of the colon leaving lesions in the right colon undetected. Both scoping procedures are expensive, require cathartic preparation and have increased risk of morbidity and mortality. Improved tests with increased sensitivity, specificity, ease of use and decreased costs are clearly needed before general widespread screening for colorectal cancer becomes routine.

Early colorectal cancer detection is generally based on the fecal occult blood test (FOBT) performed annually on asymptomatic individuals. Current recommendations adapted by several healthcare organizations, including the American Cancer Society, call for fecal occult blood testing beginning at age 50, repeated annually until such time as the patient would no longer benefit from screening. A positive FOBT leads to colonoscopic examination of the bowel; an expensive and invasive procedure, with a serious complication rate of one per 5,000 examinations. Only 12% of patients with heme positive stool are diagnosed with cancer or large polyps at the time of colonoscopy. A number of studies show that FOBT screening does not improve cancer-related mortality or overall survival. Compliance with occult blood testing has been poor; less than 20 percent of the population is offered or completes FOBT as recommended. If FOBT is properly done, the patient collects a fecal sample from three consecutive bowel movements. Samples are obtained while the patient adheres to dietary guidelines and avoids medications known to induce occult gastrointestinal bleeding. In reality, physicians frequently fail to instruct patients properly, patients frequently fail to adhere to protocol, and some patients find the task of collecting fecal samples difficult or unpleasant, hence compliance with annual occult blood testing is poor. If testing sensitivity and specificity can be improved over current methods, the frequency of testing could be reduced, collection of consecutive samples would be eliminated, dietary and medication schedule modifications would be eliminated, and patient compliance would be enhanced. Compounding the problem of compliance, the sensitivity and specificity of FOBT to detect colon cancer is poor. Poor test specificity leads to unnecessary

colonoscopy, adding considerable expense to colon cancer screening.

Specificity of the FOBT has been calculated at best to be 96%, with a sensitivity of 43% (adenomas) and 50% (colorectal carcinoma). Sensitivity can be improved using an immunoassay FOBT such as that produced under the tradename 'InSure™', with an improved sensitivity of 77 % (adenomas) and 88.9% (colorectal carcinoma.

Molecular disease markers. Molecular disease markers offer several advantages over other types of markers, one advantage being that even samples of very small sizes and/or samples whose tissue architecture has not been maintained can be analyzed quite efficiently. Within the last decade a number of genes have been shown to be differentially expressed between normal and colon carcinomas. However, no single or combination of marker has been shown to be sufficient for the diagnosis of colon carcinomas. High-dimensional mRNA based approaches have recently been shown to be able to provide a better means to distinguish between different tumor types and benign and malignant lesions. However its application as a routine diagnostic tool in a clinical environment is impeded by the extreme instability of mRNA, the rapidly occurring expression changes following certain triggers (*e.g.*, sample collection), and, most importantly, the large amount of mRNA needed for analysis (Lipshutz, R. J. et al., Nature Genetics 21:20-24, 1999; Bowtell, D. D. L. Nature genetics suppl. 21:25-32, 1999), which often cannot be obtained from a routine biopsy.

The use of biological markers to further improve sensitivity and specificity of FOBT has been suggested, examples of such tests include the PreGen-Plus™ stool analysis assay available from EXACT Sciences which has a sensitivity of 20% (adenoma) and 52% (colorectal carcinoma) and a specificity of 95% in both cases. This test assays for the presence of 23 DNA mutations associated with the development of colon neoplasms. The use of DNA methylation as colon cancer markers is known. For example Sabbioni et al. (*Molecular Diagnosis* 7:201-207, 2003) detected hypermethylation of a panel of genes consisting of TPEF, HIC1, DAPK and MGMT in peripheral blood in 98% of colon carcinoma patients. However, this does provide a suitable basis for a commercially marketable test, as the specificity of such a test must also be sufficiently high.

The current model of colorectal pathogenesis favours a stepwise progression of adenomas, which includes the development of dysplasia and finally signs of invasive cancer. The molecular changes underlying this adenoma-carcinoma sequence include genetic and epigenetic alterations of tumour suppressor genes (APC, p53, DCC), the activation of oncogenes (K-ras) and the inactivation of DNA mismatch repair genes. Recently, further molecular changes and genetic defects have been revealed. Thus, activation of the Wnt signalling pathway not only includes mutations of the APC gene, but may also result from β-catenin mutations. Furthermore, alterations in the TGF-β signalling pathway

together with its signal transducers SMAD4 and SMAD2 have been linked to the development of colon cancer.

Despite recent progress in the understanding of the pathogenesis of adenomas and carcinomas of the colon and their genetic and molecular changes, the genetic and epigenetic changes underlying the development of metastasis are less well understood. It is, however, generally well accepted that the process of invasion and proteolysis of the extracellular matrix, as well as infiltration of the vascular basement membrane involve adhesive proteins, such as members of the family of integrin receptors, the cadherins, the immunoglobulin superfamily, the laminin binding protein and the CD44 receptor. Apart from adhesion, the process of metastasis formation also includes the induction and regulation of angiogenesis (VEGF, bFGF), the induction of cell proliferation (EGF, HGF, IGF) and the activation of proteolytic enzymes (MMPs, TIMPs, uPAR), as well as the inhibition of apoptosis (Bcl-2, Bcl-X). More recently other groups have compared the genetic and molecular changes in metastatic lesions to the changes found in primary colorectal cancers. Thus, Kleeff et al. reported the loss of DOC-2, a candidate tumour suppressor gene, both in primary and metastatic colorectal cancer. Furthermore, Zauber et al. reported that in their series of 42 colorectal cancers Ki-ras mutations in the primary cancers were identical in all of the 42 paired primary and synchronous metastatic lesions. Similarly loss of heterozygosity at the APC locus was identical for 39 paired carcinomas and synchronous metastasis. The authors concluded that for Ki-ras and APC genes the genetic changes in metastasis are identical to the primary colorectal cancer. However, other groups have found genetic and molecular changes in metastatic colon cancers, that are not present in the primary cancers. Thus, the development of LOH of chromosome 3p in colorectal metastasis has been reported. In addition, using comparative genomic hybridization several alterations were found in liver metastasis that were unique to metastastic lesions (-9q, -11q, and -17q).

CpG island methylation. Apart from mutations aberrant methylation of CpG islands has been shown to lead to the transcriptional silencing of certain genes that have been previously linked to the pathogenesis of various cancers. CpG islands are short sequences which are rich in CpG dinucleotides and can usually be found in the 5' region of approximately 50% of all human genes. Methylation of the cytosines in these islands leads to the loss of gene expression and has been reported in the inactivation of the X chromosome and genomic imprinting.

Multifactorial approach. Cancer diagnostics has traditionally relied upon the detection of single molecular markers (*e.g.*, gene mutations, elevated PSA levels). Unfortunately, cancer is a disease state in which single markers have typically failed to detect or differentiate many forms of the disease. Thus, assays that recognize only a single marker have been shown to be of limited predictive value. A fundamental aspect of this invention is that methylation-based cancer diagnostics and the screening,

diagnosis, and therapeutic monitoring of such diseases will provide significant improvements over the state-of-the-art that uses single marker analyses by the use of a selection of multiple markers. The multiplexed analytical approach is particularly well suited for cancer diagnostics since cancer is not a simple disease, this multi-factorial "panel" approach is consistent with the heterogeneous nature of cancer, both cytologically and clinically.

Key to the successful implementation of a panel approach to methylation based diagnostic tests is the design and development of optimized panels of markers that can characterize and distinguish disease states. The present invention describes a plurality of particularly efficient and unique panels of genes, the methylation analysis of one or a combination of the members of the panel enabling the detection of colon cell proliferative disorders with a particularly high sensitivity, specificity and/or predictive value.

Development of medical tests. Two key evaluative measures of any medical screening or diagnostic test are its sensitivity and specificity, which measure how well the test performs to accurately detect all affected individuals without exception, and without falsely including individuals who do not have the target disease (predictive value). Historically, many diagnostic tests have been criticized due to poor sensitivity and specificity.

A true positive (TP) result is where the test is positive and the condition is present. A false positive (FP) result is where the test is positive but the condition is not present. A true negative (TN) result is where the test is negative and the condition is not present. A false negative (FN) result is where the test is negative but the condition is not present. In this context: Sensitivity = TP/(TP+FN); Specificity = TN/(FP+TN); and Predictive value = TP/(TP+FP).

Sensitivity is a measure of a test's ability to correctly detect the target disease in an individual being tested. A test having poor sensitivity produces a high rate of false negatives, *i.e.*, individuals who have the disease but are falsely identified as being free of that particular disease. The potential danger of a false negative is that the diseased individual will remain undiagnosed and untreated for some period of time, during which the disease may progress to a later stage wherein treatments, if any, may be less effective. An example of a test that has low sensitivity is a protein-based blood test for HIV. This type of test exhibits poor sensitivity because it fails to detect the presence of the virus until the disease is well established and the virus has invaded the bloodstream in substantial numbers. In contrast, an example of a test that has high sensitivity is viral-load detection using the polymerase chain reaction (PCR). High sensitivity is achieved because this type of test can detect very small quantities of the virus. High sensitivity is particularly important when the consequences of missing a diagnosis are high.

Specificity, on the other hand, is a measure of a test's ability to identify accurately patients who are free of the disease state. A test having poor specificity produces a high rate of false positives, *i.e.*, individuals who are falsely identified as having the disease. A drawback of false positives is that they force patients to undergo unnecessary medical procedures treatments with their attendant risks, emotional and financial stresses, and which could have adverse effects on the patient's health. A feature of diseases which makes it difficult to develop diagnostic tests with high specificity is that disease mechanisms, particularly in cancer, often involve a plurality of genes and proteins. Additionally, certain proteins may be elevated for reasons unrelated to a disease state. An example of a test that has high specificity is a gene-based test that can detect a p53 mutation. Specificity is important when the cost or risk associated with further diagnostic procedures or further medical intervention are very high.

Pronounced need in the art. It is generally accepted that there is a pronounced need in the art for improved screening and early detection of cancers. As an example, if colon cancer screening specificity can be increased, the problem of false positive test results leading to unnecessary colonoscopic examination would be reduced leading to cost savings and improved safety.

In view of the incidence of cancers in general and more particularly the disadvantages associated with current colorectal cell proliferative disorder screening methods there is a substantial need in the art for improved methods for the early detection of cancer, in particular colon cancer, to be used in addition to or as a substitute for currently available tests.

Background of the genes of the present invention. The human Septin 9 gene (also known as MLL septin-like fusion protein, MLL septin-like fusion protein MSF-A, Slpa, Eseptin, Msf, septin-like protein Ovarian/Breast septin (Ov/Br septin) and Septin D1) is located on chromosome 17q25 within contig AC068594.15.1.168501 and is a member of the Septin gene family. SEQ ID NO: 1 provides the sequence of said gene, comprising regions of both the Septin 9 and Q9HC74 transcripts and promoter regions.

It has been postulated that members of the Septin gene family are associated with multiple cellular functions ranging from vesicle transport to cytokinesis. Disruption of the action of *Septin 9* results in incomplete cell division, see Surka, M.C., Tsang, C.W., and Trimble, W.S. Mol Biol Cell, 13: 3532-45 (2002). *Septin 9* and other proteins have been shown to be fusion partners of the proto-oncogene *MLL* suggesting a role in tumorigenesis, see Osaka, M, Rowley, J.D. and Zeleznik-Le, N.J. PNAS, 96:6428-6433 (1999). Burrows et al. reported an in depth study of expression of the multiple isoforms of the *Septin 9* gene in ovarian cancer and showed tissue specific expression of various transcripts, see

Burrows, J. F., Chanduloy, et al. S.E.H. Journal of Pathology, 201:581-588 (2003).

A recent study of over 7000 normal and tumor tissues indicates that there is consistent over-expression of *Septin 9* isoforms in a number of tumor tissues, see Scott, M., Hyland, P.L., et al. Oncogene, 24: 4688-4700 (2005). The authors speculate that the gene is likely a type II cancer gene where changes in RNA transcript processing control regulation of different protein products, and the levels of these altered protein isoforms may provide answers to the gene's role in malignancy.

The ALX4 gene is a putative transcription factor that belongs to the family of paired-class homeoproteins. This gene is part of a family of genes that includes the mammalian genes Alx3, Cart-1, MHox, and S8 and exhibits similarity to the *Drosophila* gene *aristaless*. It binds palindromic DNA sequences (5'-TAAT-3') as either homodimers or as heterodimers with other family members and strongly activates transcription from a promoter containing the homeodomain binding site, P2. *ALX4* is expressed at several sites during development, including the craniofacial and limb-bud mesenchyme. Interestingly, *ALX4* deficient mice exhibit body-wall defects, preaxial polydactyly, and a decreased size of the parietal plate of the skull, while mutations of the human homeobox gene *ALX4* have been found in inherited defects of skull ossification. *ALX4* is also expressed in various tissues whose development is dependent on epithelial-mesenchymal interactions and regulates mesenchymal-specific activities of LEF-1.

Methylation of the gene ALX4 has been previously disclosed in WO 2004/035803. In said document it was disclosed that ALX4 was methylated accross all classes of colon adenoma and polyp classes, with no differentation between benign, malignant and pre-malignant classes thereof. The subject matter of the present invention differs from that of WO 2004/035803 in that the method of the present invention is practised on body fluid isolated from a subject. The technical effect of practising the method on body fluid samples is that it enables the differentiation between benign and pre-cancerous lesions. Thus the technical problem to be solved by the present invention is how to differentiate between harmless (i.e. benign) and potentially harmful (i.e. those undergoing malignant transformation) colorectal lesions. The person of skill in the art when taking the teachings of WO 2004/035803 into account would not be led to analyse the methylation of said markers in body fluids as opposed to e.g. a histological sample.

SUMMARY OF THE INVENTION

The present invention provides a method for determining the presence or absence of pre-cancerous colorectal lesions in a subject comprising determining the expression levels of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4 in a body fluid sample isolated from said subject wherein underexpression and/or

CpG methylation is indicative of the presence of said lesions. Alternatively, said invention provides a method for the differentiation of pre-cancerous from benign colorectal lesions. Various aspects of the present invention provide an efficient and unique genetic marker, whereby expression analysis of said marker enables the detection of pre-cancerous lesions with a particularly high sensitivity, specificity and/or predictive value. The inventive testing methods have particular utility for the screening of at-risk populations. The inventive methods have advantages over prior art methods (including the industry standard FOBT) because it enables the detection of colorectal lesions undergoing malignant transformation but prior to the development of cancer. Furthermore the high sensitivity and specificity as well as non-invasiveness of such a test is likely to result in increased patient compliance.

In one embodiment the invention provides a method for detecting and/or classifying cell proliferative disorders in a subject comprising determining the expression levels of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4 in a body fluid sample isolated from said subject wherein under-expression and/or CpG methylation is indicative of the presence or class of said disorder. In one embodiment said expression level is determined by detecting the presence, absence or level of mRNA transcribed from said gene. In a further embodiment said expression level is determined by detecting the presence, absence or level of a polypeptide encoded by said gene or sequence thereof.

In a further preferred embodiment said expression is determined by detecting the presence or absence of CpG methylation within said gene(s), wherein the presence of methylation indicates the presence of a cell proliferative lesion undergoing or having already achieved malignant transformation. Said method comprises the following steps: i) contacting genomic DNA isolated from a body fluid sample (preferably selected from the group consisting of blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood) 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 target region of the genomic DNA, wherein the nucleotide sequence of said target region comprises at least one CpG dinucleotide sequence of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4; and ii) detecting and/or classifying cell proliferative disorders , at least in part. Preferably the target region comprises, or hybridizes under stringent conditions to a sequence of at least 16 contiguous nucleotides of at least one sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2.

The method is novel as no methods currently exist that enable the early detection of potentially harmful colorectal lesions. For example, current methods used to detect and diagnose colorectal carcinoma include colonoscopy, sigmoidoscopy, and fecal occult blood colon cancer. In comparison to these methods, the disclosed invention is much less invasive than colonoscopy, and as, if not more

sensitive than sigmoidoscopy and FOBT while enabling the detection of harmful lesions before reaching the carcinoma stage. The development of a body fluid assay represents a clear technical advantage over current methods known in the art in that it is anticipated that at least for colorectal carcinoma screening patient compliance for a single body fluid based test will be higher than the triplicate analysis of stool currently recommended for FOBT.

A particular embodiment of the method comprises the use of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4 as a marker for the detection and/or classification of colorectal cellular proliferative disorders. The present invention is particularly suited for the detection of pre-cancerous colorectal cellular proliferative disorders undergoing malignant transformation. Said use of the gene may be enabled by means of any analysis of the expression of the gene, by means of mRNA expression analysis or protein expression analysis. However, in the most preferred embodiment of the invention, the detection, differentiation and distinguishing of colorectal cell proliferative disorders is enabled by means of analysis of the *methylation status* of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4 and their promoter or regulatory elements.

The invention provides a method for the analysis of biological samples for features associated with the development of pre-cancerous cellular proliferative disorders, the method characterized in that at least one nucleic acid, or a fragment thereof, from the group consisting of SEQ ID NO: 1 TO SEQ ID NO: 2 is contacted with a reagent or series of reagents capable of distinguishing between methylated and non methylated CpG dinucleotides within the genomic sequence, or sequences of interest.

The present invention provides a method for ascertaining epigenetic parameters of genomic DNA associated with the development of malignant colorectal cellular proliferative disorders, the method has utility for the improved diagnosis, treatment and monitoring of said diseases.

Preferably, the source of the test sample is selected from the group consisting of cells or cell lines, histological slides, biopsies, paraffin-embedded tissue, body fluids, stool, urine, blood, and combinations thereof. More preferably, the source is selected from the group consisting of stool, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood obtained from the subject.

Specifically, the present invention provides a method for detecting pre-cancerous colorectal cellular proliferative disorders or for differentiating between pre-cancerous and benign cellular proliferative disorders, comprising: obtaining a body fluid sample comprising genomic nucleic acid(s); contacting

the nucleic acid(s), or a fragment thereof, with one reagent or a plurality of reagents sufficient for distinguishing between methylated and non methylated CpG dinucleotide sequences within at least one target sequence of the subject nucleic acid, wherein the target sequence comprises, or hybridises under stringent conditions to, a sequence comprising at least 16 contiguous nucleotides of a sequence selected from the group consisting SEQ ID NO: 1 TO SEQ ID NO: 2, said contiguous nucleotides comprising at least one CpG dinucleotide sequence; and determining, based at least in part on said distinguishing, the methylation state of at least one target CpG dinucleotide sequence, or an average, or a value reflecting an average methylation state of a plurality of target CpG dinucleotide sequences.

Preferably, distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises methylation state-dependent conversion or non-conversion of at least one such CpG dinucleotide sequence to the corresponding converted or non-converted dinucleotide sequence within a sequence selected from the group consisting of SEQ ID NO: 3 to SEQ ID NO: 10, and contiguous regions thereof corresponding to the target sequence.

Further embodiments provide alternative methods comprising: obtaining a body fluid sample having subject genomic DNA; extracting the genomic DNA; contacting the genomic DNA, or a fragment thereof, comprising one or more sequences selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2 or a sequence that hybridizes under stringent conditions thereto, with one or more methylation-sensitive restriction enzymes, wherein the genomic DNA is either digested thereby to produce digestion fragments, or is not digested thereby; and determining, based on a presence or absence of, or on property of at least one such fragment, the methylation state of at least one CpG dinucleotide sequence of at least one genomic sequence selected form the group consisting of SEQ ID NO: 1 TO SEQ ID NO: 2, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences thereof. Preferably, the digested or undigested genomic DNA is amplified prior to said determining.

Additional embodiments provide genomic and chemically modified nucleic acid sequences, as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine methylation patterns within sequences from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2.

DETAILED DESCRIPTION OF THE INVENTION

Definitions:

The term "Observed/Expected Ratio" ("O/E Ratio") refers to the frequency of CpG dinucleotides within a particular DNA sequence, and corresponds to the [number of CpG sites / (number of C bases x number of G bases)] / band length for each fragment.

The term "CpG island" refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an "Observed/Expected Ratio" >0.6, and (2) having a "GC Content" >0.5. CpG islands are typically, but not always, between about 0.2 to about 1 KB, or to about 2kb in length.

The term "methylation state" or "methylation status" refers to the presence or absence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence. Methylation states at one or more particular CpG methylation sites (each having two CpG dinucleotide sequences) within a DNA sequence include "unmethylated," "fully-methylated" and "hemi-methylated."

The term "hemi-methylation" or "hemimethylation" refers to the methylation state of a double stranded DNA wherein only one strand thereof is methylated.

The term 'AUC' as used herein is an abbreviation for the area under a curve. In particular it refers to the area under a Receiver Operating Characteristic (ROC) curve. The ROC curve is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test. It shows the trade-off between sensitivity and specificity depending on the selected cut point (any increase in sensitivity will be accompanied by a decrease in specificity). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test (the larger the area the better, 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. Signal Detection Theory and ROC Analysis, Academic Press, New York, 1975).

The term "hypermethylation" refers to the average methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

The term "hypomethylation" refers to the average methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

The term "microarray" refers broadly to both "DNA microarrays," and 'DNA chip(s),' as recognized in the art, encompasses all art-recognized solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

"Epigenetic parameters" are, in particular, cytosine methylation. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analysed using the described method but which, in turn, correlate with the DNA methylation.

The term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

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

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., *Cancer Research* 57:594-599, 1997.

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

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

The term "HeavyMethyl™ MethyLight™" assay, in the embodiment thereof implemented herein, 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 and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

The term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described

by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

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

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

The term "hybridization" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

"Stringent hybridization conditions," as defined herein, involve hybridizing at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognized equivalent thereof (e.g., conditions in which a hybridization is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognized equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley and Sons, N.Y.) at Unit 2.10.

The terms "Methylation-specific restriction enzymes" or "methylation-sensitive restriction enzymes" shall be taken to mean an enzyme that selectively digests a nucleic acid dependant on the methylation state of its recognition site. In the case of such restriction enzymes which specifically cut if the recognition site is not methylated or hemimethylated, the cut will not take place, or with a significantly reduced efficiency, if the recognition site is methylated. In the case of such restriction enzymes which specifically cut if the recognition site is methylated, the cut will not take place, or with a significantly reduced efficiency if the recognition site is not methylated. Preferred are methylation-specific restriction enzymes, the recognition sequence of which contains a CG dinucleotide (for instance 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.

"Non-methylation-specific restriction enzymes" or "non-methylation-sensitive restriction enzymes" are restriction enzymes that cut a nucleic acid sequence irrespective of the methylation state with nearly identical efficiency. They are also called "methylation-unspecific restriction enzymes."

The term "gene" shall be taken to include all transcript variants thereof (e.g. the term "Septin 9" shall include for example its truncated transcript Q9HC74) and all promoter and regulatory elements thereof. Furthermore as a plurality of SNPs are known within said gene the term shall be taken to include all sequence variants thereof.

Colorectal lesions which do not have malignant potential are histologically classified as benign and include hyperplastic polyps, hamartomas and inflammatory polyps. Colorectal lesions which do have malignant potential are histologically classified as neoplastic adenomas and include tubular adenomas (0%-25% villious tissue), tubuvillous adenomas (25%-75% villious tissue) and villous adenomas (75%-100% villious tissue).

The terms "pre-cancerous" or "pre-malignant" and equivalents thereof shall be taken to mean any cellular proliferative disorder which is undergoing malignant transformation, such as but not limited to neoplastic adenomas including those described above. Examples of such conditions include, in the context of colorectal cellular proliferative disorders, cellular proliferative disorders (such as but not limited to those commonly referred to as "polyps") with a high degree of dysplasia and the following classes of adenomas:
Level 1: penetration of malignant glands through the muscularis mucosa into the submucosa, within the polyp head;
Level 2: the same submucosal invasion, but present at the junction of the head to the stalk;
Level 3: invasion of the stalk; and
Level 4: invasion of the stalk's base at the connection to the colonic wall (this level corresponds to stage Dukes A).

Overview:
The present invention provides a method for detecting pre-cancerous colorectal cell proliferative disorders (e.g. neoplastic adenomas) or for differentiating between benign colorectal lesions and pre-malignant colorectal lesions in a subject comprising determining the expression levels of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4 in a biological sample isolated from said subject wherein underexpression and/or CpG methylation is indicative of the presence of said disorder. Preferably the expression levels of both Septin 9 (including any transcript variants thereof) and ALX4 are analysed. Said markers may be used for the early detection of colorectal cancers during the pre-cancerous stages of the disease both by detecting the presence thereof and/or differentiating between benign and malignant forms of colorectal lesions.

In a first embodiment the present invention is based upon the analysis of CpG methylation status of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4. Preferably the methylation status of both Septin 9 (including any transcript variants thereof) and ALX4 are analyzed. It is further preferred that the sequences of said genes are as according to TABLE 1.

Bisulfite modification of DNA is an art-recognized tool used to assess CpG methylation status. 5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing, because 5-methylcytosine has the same base pairing behavior as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during, *e.g.*, PCR amplification.

The most frequently used method for analyzing DNA for the presence of 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine whereby, upon subsequent alkaline hydrolysis, cytosine is converted to uracil which corresponds to thymine in its base pairing behavior. Significantly, however, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is *converted* in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridization behavior, can now be detected as the only remaining cytosine using standard, art-recognized molecular biological techniques, for example, by amplification and hybridization, or by sequencing. All of these techniques are based on differential base pairing properties, which can now be fully exploited.

The prior art, in terms of sensitivity, is defined by a method 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 all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of art-recognized methods for detecting 5-methylcytosine is provided by Rein, T., et al., *Nucleic Acids Res.*, 26:2255, 1998.

The bisulfite technique, barring few exceptions (*e.g.*, Zeschnigk M, et al., *Eur J Hum Genet.* 5:94-98, 1997), is currently only used in research. In all instances, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment, and either completely sequenced (Olek and Walter, *Nat Genet.* 1997 17:275-6, 1997), subjected to one or more primer extension reactions (Gonzalgo and

Jones, *Nucleic Acids Res.*, 25:2529-31, 1997; WO 95/00669; U.S. Patent No. 6,251,594) to analyze individual cytosine positions, or treated by enzymatic digestion (Xiong and Laird, *Nucleic Acids Res.*, 25:2532-4, 1997). 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 and Clark, *Bioessays*, 16:431-6, 1994; Zeschnigk M, et al., *Hum Mol Genet.*, 6:387-95, 1997; Feil R, et al., *Nucleic Acids Res.*, 22:695-, 1994; Martin V, et al., *Gene*, 157:261-4, 1995; WO 9746705 and WO 9515373).

The present invention provides for the use of the bisulfite technique, in combination with one or more methylation assays, for determination of the methylation status of CpG dinucleotide sequences within at least one sequence selected from the group consisting SEQ ID NO: 1 TO SEQ ID NO: 2. It is particularly preferred that CpG positions of the gene ALX4 within bases 42,700-52,000 of SEQ ID NO: 1 (or said positions within the equivalent bisulfite converted sequences) are analyzed. It is particularly preferred that CpG positions of the gene Septin 9 within bases 1,000-4,250 or 93,850-96,000 of SEQ ID NO: 2 (or said positions within the equivalent bisulfite converted sequences) are analysed. 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 colorectal cells within a background of blood or stool. Accordingly, when analyzing the methylation status of a CpG position within such a sample the person skilled in the art may use a quantitative assay for determining the level (*e.g.*, percent, fraction, ratio, proportion or degree) of methylation at a particular CpG position as opposed to a methylation state. Accordingly the term methylation status or methylation state should also be taken to mean a value reflecting the degree of methylation at a CpG position. Unless specifically stated the terms "hypermethylated" or "upmethylated" shall be taken to mean a methylation level above that of a specified cut-off point, wherein said cut-off may be a value representing the average or median methylation level for a given population, or is preferably an optimized cut-off level. The "cut-off" is also referred herein as a "threshold". In the context of the present invention the terms "methylated", "hypermethylated" or "upmethylated" shall be taken to include a methylation level above the cut-off be zero (0) % (or equivalents thereof) methylation for all CpG positions within and associated with (e.g. in promoter or regulatory regions) the genes selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4.

According to the present invention, determination of the methylation status of CpG dinucleotide sequences within SEQ ID NO: 1 TO SEQ ID NO: 2 has utility in the detection of colorectal lesions undergoing malignant transformation and thus in the early detection of colorectal cancers.

Methylation Assay Procedures. Various methylation assay procedures are known in the art, and can be

used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (*e.g.*, CpG islands) within a DNA sequence. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, 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 DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, *e.g.*, the method described by Sadri and Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong and Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

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

Typical reagents (*e.g.*, as might be found in a typical COBRA™-based kit) for COBRA™ analysis may include, but are not limited to: PCR primers for specific gene (or bisulfite treated DNA sequence or CpG island); 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 and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent 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 other 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 (also referred to herein as *blockers*) covering CpG positions between, or covered by the amplification primers enable methylation-specific selective amplification of a nucleic acid sample.

The term "HeavyMethyl™ MethyLight™" assay, in the embodiment thereof implemented herein, 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 genes (or bisulfite treated DNA sequence or CpG island); blocking oligonucleotides; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

MSP. 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; US Patent No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all unmethylated, but not methylated cytosines to uracil, and 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 gene (or bisulfite treated DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

MethyLight™. The MethyLight™ assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqMan®) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 59:2302-2306, 1999). Briefly, the MethyLight™ process begins with a mixed sample of genomic DNA that is converted, in a sodium

bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur both at the level of the amplification process and at the level of the fluorescence detection process.

The MethyLight™ assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for a methylation specific amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of 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 MethyLight™ process can by used with any suitable probes e.g. "TaqMan®" , Lightcycler® etc.... For example, 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 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 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

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

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

fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of 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 by used with any suitable probes e.g. "TaqMan®", Lightcycler® etc... 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 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system. Typical reagents (*e.g.*, as might be found in a typical QM™ -based kit) for QM™ analysis may include, but are not limited to: PCR primers for specific gene (or bisulfite treated DNA sequence or CpG island); TaqMan® or Lightcycler® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

Ms-SNuPE. 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 and 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(s) of interest. Small amounts of DNA can be analyzed (*e.g.*, micro-dissected 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 gene (or bisulfite treated DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE™ primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and labelled nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation

buffer; sulfonation buffer; DNA recovery regents or kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

The genomic sequence according to SEQ ID NO: 1 to SEQ ID NO: 2, and non-naturally occurring treated variants thereof according to SEQ ID NO: 3 to SEQ ID NO: 10, were determined to have novel utility for the detection of malignant or pre-malignant colorectal adenomas and the differentiation of benign colorectal lesions from those undergoing malignant transformation and accordingly to be of use in the early detection of colorectal carcinomas.

In one embodiment the invention of the method comprises the following steps: i) contacting genomic DNA (preferably isolated from body fluids) 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 gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4 (including their promoter and regulatory regions); and ii) determining the presence or absence of pre-cancerous (i.e. malignant or pre-malignant) colon cellular proliferative disorders or distinguishing between or among benign and pre-cancerous colon cellular proliferative disorders. Preferably the methylation status of both Septin 9 (including any transcript variants thereof) and ALX4 are analyzed.

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

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 target region of the genomic DNA, wherein the target region comprises, or hybridizes under stringent conditions to a sequence of at least 16 contiguous nucleotides of at least one sequence selected from

the group consisting of SEQ ID NO: 1 TO SEQ ID NO: 2 respectively, wherein said contiguous nucleotides comprise at least one CpG dinucleotide sequence.

It is particularly preferred that said 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 an alternative embodiment said reagent may be a methylation sensitive restriction enzyme.

Wherein the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior It is preferred that this treatment is carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis. Such a treatment results in the conversion of SEQ ID NO: 1 TO SEQ ID NO: 2 to SEQ ID NO: 3-6, (respectively) wherein said CpG dinucleotides are methylated or SEQ ID NO: 7-10 wherein said CpG dinucleotides are unmethylated.

The treated DNA is then analyzed in order to determine the methylation state of the target gene sequences (at least one gene or genomic sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2). It is particularly preferred that the target region comprises, or hybridizes under stringent conditions to at least 16 contiguous nucleotides of at least one gene or genomic sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2. It is preferred that the sequence of said genes according to SEQ ID NO: 1 to SEQ ID NO: 2 are analyzed. The method of analysis may be selected from those known in the art, including those listed herein. Particularly preferred are MethyLight™, MSP and the use of blocking oligonucleotides (HeavyMethyl™) as described herein. It is further preferred that any oligonucleotides used in such analysis (including primers, blocking oligonucleotides and detection probes) should be reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one or more of SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto.

Aberrant methylation, more specifically hypermethylation of the genes and genomic sequences thereof according to Table 1 selected from the group consisting of Septin 9 (including all transcript variants thereof) and/or ALX4 (including their promoter and/or regulatory regions) is associated with the presence of cancer. Preferably the methylation status of both Septin 9 (including any transcript variants thereof) and ALX4 are analyzed. Accordingly wherein a biological sample presents within any degree of methylation, said sample should be determined as having malignant potential .

The method of the invention may alternatively be enabled by means of any analysis of the expression of an RNA transcribed therefrom or polypeptide or protein translated from said RNA, preferably by means of mRNA expression analysis or polypeptide expression analysis. Accordingly the present invention also provides assays and methods, both quantitative and qualitative for detecting the expression of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and/or ALX4 in a subject and determining therefrom upon the presence or absence of a neoplastic colorectal cell proliferative disorder, or differentiating between a malignant or pre-malignant and benign colorectal cell proliferative disorder.

Aberrant expression of mRNA transcribed from the genes or genomic sequences selected from the group consisting of Septin 9 (including all transcript variants thereof) and/or ALX4 are associated with the malignant transformation of colorectal lesions. Preferably the expression of both Septin 9 (including any transcript variants thereof) and ALX4 are analyzed. According to the present invention, under expression (and/or methylation) is associated with the presence of malignant or pre-malignant colorectal cellular proliferative disorders, and over-expression (and/or absence of methylation) is associated with benign colorectal cellular proliferative disorders. It is particularly preferred that the expression of at least one of the transcript variants of the genes Septin 9 and ALX4 is determined.

To detect the presence of mRNA encoding a gene or genomic sequence, a sample is obtained from a patient. The sample may be any suitable sample comprising cellular matter of the lesion. Suitable sample types include cell lines, histological slides, biopsies, paraffin-embedded tissue, body fluids, stool, colonic effluent, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood and all possible combinations thereof. It is preferred that said sample types are stool or body fluids selected from the group consisting colonic effluent, urine, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood.

The sample may be treated to extract the RNA contained therein. The resulting nucleic acid from the sample is then analyzed. Many techniques are known in the state of the art for determining absolute and relative levels of gene expression, commonly used techniques suitable for use in the present invention include in situ hybridization (*e.g.*, FISH), Northern analysis, RNase protection assays (RPA), microarrays and PCR-based techniques, such as quantitative PCR and differential display PCR or any other nucleic acid detection method.

Particularly preferred is the use of the reverse transcription/polymerization chain reaction technique (RT-PCR). The method of RT-PCR is well known in the art (for example, see Watson and Fleming, supra).

The RT-PCR method can be performed as follows. Total cellular RNA is isolated by, for example, the standard guanidium isothiocyanate method and the total RNA is reverse transcribed. The reverse transcription method involves synthesis of DNA on a template of RNA using a reverse transcriptase enzyme and a 3' end oligonucleotide dT primer and/or random hexamer primers. The cDNA thus produced is then amplified by means of PCR. (Belyavsky et al, Nucl Acid Res 17:2919-2932, 1989; Krug and Berger, Methods in Enzymology, Academic Press, N.Y., Vol.152, pp. 316-325, 1987 which are incorporated by reference). Further preferred is the "Real-time" variant of RT- PCR, wherein the PCR product is detected by means of hybridization probes (e.g. TaqMan, Lightcycler, Molecular Beacons and Scorpion) or SYBR green. The detected signal from the probes or SYBR green is then quantitated either by reference to a standard curve or by comparing the Ct values to that of a calibration standard. Analysis of housekeeping genes is often used to normalize the results.

In Northern blot analysis total or poly(A)+ mRNA is run on a denaturing agarose gel and detected by hybridisation to a labelled probe in the dried gel itself or on a membrane. The resulting signal is proportional to the amount of target RNA in the RNA population.

Comparing the signals from two or more cell populations or tissues reveals relative differences in gene expression levels. Absolute quantitation can be performed by comparing the signal to a standard curve generated using known amounts of an in vitro transcript corresponding to the target RNA. Analysis of housekeeping genes, genes whose expression levels are expected to remain relatively constant regardless of conditions, is often used to normalize the results, eliminating any apparent differences caused by unequal transfer of RNA to the membrane or unequal loading of RNA on the gel.

The first step in Northern analysis is isolating pure, intact RNA from the cells or tissue of interest. Because Northern blots distinguish RNAs by size, sample integrity influences the degree to which a signal is localized in a single band. Partially degraded RNA samples will result in the signal being smeared or distributed over several bands with an overall loss in sensitivity and possibly an erroneous interpretation of the data. In Northern blot analysis, DNA, RNA and oligonucleotide probes can be used and these probes are preferably labelled (*e.g.*, radioactive labels, mass labels or fluorescent labels). The size of the target RNA, not the probe, will determine the size of the detected band, so methods such as random-primed labelling, which generates probes of variable lengths, are suitable for probe synthesis. The specific activity of the probe will determine the level of sensitivity, so it is preferred that probes with high specific activities, are used.

In an RNase protection assay, the RNA target and an RNA probe of a defined length are hybridised in solution. Following hybridisation, the RNA is digested with RNases specific for single-stranded nucleic acids to remove any unhybridized, single-stranded target RNA and probe. The RNases are

inactivated, and the RNA is separated e.g. by denaturing polyacrylamide gel electrophoresis. The amount of intact RNA probe is proportional to the amount of target RNA in the RNA population. RPA can be used for relative and absolute quantitation of gene expression and also for mapping RNA structure, such as intron/exon boundaries and transcription start sites. The RNase protection assay is preferable to Northern blot analysis as it generally has a lower limit of detection.

The antisense RNA probes used in RPA are generated by in vitro transcription of a DNA template with a defined endpoint and are typically in the range of 50–600 nucleotides. The use of RNA probes that include additional sequences not homologous to the target RNA allows the protected fragment to be distinguished from the full-length probe. RNA probes are typically used instead of DNA probes due to the ease of generating single-stranded RNA probes and the reproducibility and reliability of RNA:RNA duplex digestion with RNases (Ausubel *et al.* 2003), particularly preferred are probes with high specific activities.

Particularly preferred is the use of microarrays. The microarray analysis process can be divided into two main parts. First is the immobilization of known gene sequences onto glass slides or other solid support followed by hybridisation of the fluorescently labelled cDNA (comprising the sequences to be interrogated) to the known genes immobilized on the glass slide (or other solid phase). After hybridisation, arrays are scanned using a fluorescent microarray scanner. Analysing the relative fluorescent intensity of different genes provides a measure of the differences in gene expression.

DNA arrays can be generated by immobilizing presynthesized oligonucleotides onto prepared glass slides or other solid surfaces. In this case, representative gene sequences are manufactured and prepared using standard oligonucleotide synthesis and purification methods. These synthesized gene sequences are complementary to the RNA transcript(s) of the genes of interest (in this case the genes or genomic sequences selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4) and tend to be shorter sequences in the range of 25–70 nucleotides. In a preferred embodiment said oligonucleotides or polynucleotides comprise at least 9, 18 or 25 bases of a sequence complementary to or hybridising to at least the mRNA transcript and sequences complementary thereto. Alternatively, immobilized oligos can be chemically synthesized in situ on the surface of the slide. *In si*tu oligonucleotide synthesis involves the consecutive addition of the appropriate nucleotides to the spots on the microarray; spots not receiving a nucleotide are protected during each stage of the process using physical or virtual masks. Preferably said synthesized nucleic acids are locked nucleic acids.

In expression profiling microarray experiments, the RNA templates used are representative of the transcription profile of the cells or tissues under study. RNA is first isolated from the cell populations

or tissues to be compared. Each RNA sample is then used as a template to generate fluorescently labelled cDNA via a reverse transcription reaction. Fluorescent labelling of the cDNA can be accomplished by either direct labelling or indirect labelling methods. During direct labelling, fluorescently modified nucleotides (*e.g.*, Cy®3- or Cy®5-dCTP) are incorporated directly into the cDNA during the reverse transcription. Alternatively, indirect labelling can be achieved by incorporating aminoallyl-modified nucleotides during cDNA synthesis and then conjugating an N-hydroxysuccinimide (NHS)-ester dye to the aminoallyl-modified cDNA after the reverse transcription reaction is complete. Alternatively, the probe may be unlabelled, but may be detectable by specific binding with a ligand which is labelled, either directly or indirectly. Suitable labels and methods for labelling ligands (and probes) are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (*e.g.*, nick translation or kinasing). Other suitable labels include but are not limited to biotin, fluorescent groups, chemiluminescent groups (*e.g.*, dioxetanes, particularly triggered dioxetanes), enzymes, antibodies, and the like.

To perform differential gene expression analysis, cDNA generated from different RNA samples are labelled with Cy®3. The resulting labelled cDNA is purified to remove unincorporated nucleotides, free dye and residual RNA. Following purification, the labelled cDNA samples are hybridised to the microarray. The stringency of hybridisation is determined by a number of factors during hybridisation and during the washing procedure, including temperature, ionic strength, length of time and concentration of formamide. These factors are outlined in, for example, Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd ed., 1989). The microarray is scanned post-hybridisation using a fluorescent microarray scanner. The fluorescent intensity of each spot indicates the level of expression of the analysed gene; bright spots correspond to strongly expressed genes, while dim spots indicate weak expression.

Once the images are obtained, the raw data must be analysed. First, the background fluorescence must be subtracted from the fluorescence of each spot. The data is then normalized to a control sequence, such as exogenously added nucleic acids (preferably RNA or DNA), or a housekeeping gene panel to account for any non-specific hybridisation, array imperfections or variability in the array set-up, cDNA labelling, hybridisation or washing. Data normalization allows the results of multiple arrays to be compared.

Another aspect of the invention relates to a kit for use in the detection of pre-cancerous colorectal cellular proliferative disorders or the differentiation between malignant/pre-malignant and benign colorectal lesions in a subject according to the methods of the present invention, said kit comprising: a means for measuring the level of transcription of genes or genomic sequences selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4. In a preferred embodiment

the means for measuring the level of transcription comprise oligonucleotides or polynucleotides able to hybridise under stringent or moderately stringent conditions to the transcription products of a gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4. Preferably said oligonucleotides or polynucleotides are able to hybridise under stringent or moderately stringent conditions to at least one of the transcription products of a gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and/or ALX4. In one embodiment said oligonucleotides or polynucleotides comprise at least 9, 18 or 25 bases of a sequence complementary to or hybridising to at least one of said sequence or sequences complementary thereto.

In a most preferred embodiment the level of transcription is determined by techniques selected from the group of Northern Blot analysis, reverse transcriptase PCR, real-time PCR, RNAse protection, and microarray. In another embodiment of the invention the kit further comprises means for obtaining a biological sample of the patient. Preferred is a kit, which further comprises a container which is most preferably suitable for containing the means for measuring the level of transcription and the biological sample of the patient, and most preferably further comprises instructions for use and interpretation of the kit results.

In a preferred embodiment the kit comprises (a) a plurality of oligonucleotides or polynucleotides able to hybridise under stringent or moderately stringent conditions to the transcription products of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4; (b) a container, preferably suitable for containing the oligonucleotides or polynucleotides and a biological sample of the patient comprising the transcription products wherein the oligonucleotides or polynucleotides can hybridise under stringent or moderately stringent conditions to the transcription products, (c) means to detect the hybridisation of (b); and optionally, (d) instructions for use and interpretation of the kit results. It is further preferred that said oligonucleotides or polynucleotides of (a) comprise in each case at least 9, 18 or 25 bases of a sequence complementary to or hybridising to the transcription products and sequences complementary thereto.

The kit may also contain other components such as hybridisation buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimised for primer extension mediated by the polymerase, such as PCR. Preferably said polymerase is a reverse transcriptase. It is further preferred that said kit further contains an Rnase reagent.

The present invention further provides for methods for the detection of the presence of the polypeptide

encoded by said gene sequences in a sample obtained from a patient.

Aberrant levels of polypeptide expression of the polypeptides encoded by the genes or genomic sequences selected from the group consisting of Septin 9 (including all transcript variants thereof) and/or ALX4 are associated with the presence of colorectal cellular proliferative disorders. Furthermore said aberrant levels of expression are of use in the differentiation between benign and malignant/pre-malignant colorectal lesions. According to the present invention, under expression of said polypeptides is associated with the presence of colorectal lesions undergoing malignant transformation.

Any method known in the art for detecting polypeptides can be used. Such methods include, but are not limited to mass-spectrometry, immunodiffusion, immunoelectrophoresis, immunochemical methods, binder-ligand assays, immunohistochemical techniques, agglutination and complement assays (*e.g.*, see Basic and Clinical Immunology, Sites and Terr, eds., Appleton and Lange, Norwalk, Conn. pp 217-262, 1991 which is incorporated by reference). Preferred are binder-ligand immunoassay methods including reacting antibodies with an epitope or epitopes and competitively displacing a labelled polypeptide or derivative thereof.

Certain embodiments of the present invention comprise the use of antibodies specific to the polypeptide encoded by a gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4.

Such antibodies are useful for the analysis of colorectal lesions. In certain embodiments production of monoclonal or polyclonal antibodies can be induced by the use of an epitope encoded by a polypeptide of the genes Septin 9 (including all transcript variants thereof) and/or ALX4 as an antigene. Such antibodies may in turn be used to detect expressed polypeptides as markers for the early detection of colorectal cancer. The levels of such polypeptides present may be quantified by conventional methods. Antibody-polypeptide binding may be detected and quantified by a variety of means known in the art, such as labelling with fluorescent or radioactive ligands. The invention further comprises kits for performing the above-mentioned procedures, wherein such kits contain antibodies specific for the investigated polypeptides.

Numerous competitive and non-competitive polypeptide binding immunoassays are well known in the art. Antibodies employed in such assays may be unlabelled, for example as used in agglutination tests, or labelled for use a wide variety of assay methods. Labels that can be used include radionuclides, enzymes, fluorescers, chemiluminescers, enzyme substrates or co-factors, enzyme inhibitors, particles, dyes and the like. Preferred assays *include but are not limited to radioimmunoassay (RIA), enzyme*

immunoassays, e.g., enzyme-linked immunosorbent assay (ELISA), fluorescent immunoassays and the like. Polyclonal or monoclonal antibodies or epitopes thereof can be made for use in immunoassays by any of a number of methods known in the art.

In an alternative embodiment of the method the proteins may be detected by means of western blot analysis. Said analysis is standard in the art, briefly proteins are separated by means of electrophoresis, *e.g.*, SDS-PAGE. The separated proteins are then transferred to a suitable membrane (or paper), *e.g.*, nitrocellulose, retaining the spatial separation achieved by electrophoresis. The membrane is then incubated with a blocking agent to bind remaining sticky places on the membrane, commonly used agents include generic protein (*e.g.*, milk protein). An antibody specific to the protein of interest is then added, said antibody being detectably labelled for example by dyes or enzymatic means (*e.g.*, alkaline phosphatase or horseradish peroxidase). The location of the antibody on the membrane is then detected.

In an alternative embodiment of the method the proteins may be detected by means of immunohistochemistry (the use of antibodies to probe specific antigens in a sample). Said analysis is standard in the art, wherein detection of antigens in tissues is known as immunohistochemistry, while detection in cultured cells is generally termed immunocytochemistry. Briefly, the primary antibody to be detected by binding to its specific antigen. The antibody-antigen complex is then bound by a secondary enzyme conjugated antibody. In the presence of the necessary substrate and chromogen the bound enzyme is detected according to coloured deposits at the antibody-antigen binding sites. There is a wide range of suitable sample types, antigen-antibody affinity, antibody types, and detection enhancement methods. Thus optimal conditions for immunohistochemical or immunocytochemical detection must be determined by the person skilled in the art for each individual case.

One approach for preparing antibodies to a polypeptide is the selection and preparation of an amino acid sequence of all or part of the polypeptide, chemically synthesising the amino acid sequence and injecting it into an appropriate animal, usually a rabbit or a mouse (Milstein and Kohler Nature 256:495-497, 1975; Gulfre and Milstein, Methods in Enzymology: Immunochemical Techniques 73:1-46, Langone and Banatis eds., Academic Press, 1981 which are incorporated by reference in its entirety). Methods for preparation of the polypeptides or epitopes thereof include, but are not limited to chemical synthesis, recombinant DNA techniques or isolation from biological samples.

In the final step of the method the diagnosis of the patient is determined, whereby under-expression (of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4) is indicative of the presence of a colorectal lesion undergoing malignant transformation. Preferably the expression of both Septin 9 (including any transcript variants

thereof) and ALX4 are analyzed. The term under-expression shall be taken to mean expression at a detected level less than a pre-determined cut off which may be selected from the group consisting of the mean, median or an optimised threshold value.

Another aspect of the invention provides a kit for use in the early detection of colorectal cancer and/or differentiation between malignant or pre-malignant and benign colorectal lesions in a subject according to the methods of the present invention, comprising: a means for detecting polypeptides at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4. The means for detecting the polypeptides comprise preferably antibodies, antibody derivatives, or antibody fragments. The polypeptides are most preferably detected by means of Western Blotting utilizing a labelled antibody. In another embodiment of the invention the kit further comprising means for obtaining a biological sample of the patient. Preferred is a kit, which further comprises a container suitable for containing the means for detecting the polypeptides in the biological sample of the patient, and most preferably further comprises instructions for use and interpretation of the kit results. In a preferred embodiment the kit comprises: (a) a means for detecting polypeptides at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4; (b) a container suitable for containing the said means and the biological sample of the patient comprising the polypeptides wherein the means can form complexes with the polypeptides; (c) a means to detect the complexes of (b); and optionally (d) instructions for use and interpretation of the kit results. The kit may also contain other components such as buffers or solutions suitable for blocking, washing or coating, packaged in a separate container.

Particular embodiments of the present invention provide a novel application of the analysis of methylation levels and/or patterns within said sequences that enables the early detection of colorectal cancers. Early detection of cancer is directly linked with disease prognosis, and the disclosed method thereby enables the physician and patient to make better and more informed treatment decisions.

FURTHER IMPROVEMENTS

The present invention provides novel uses for the genomic sequence SEQ ID NO: 1 to SEQ ID NO: 2. Additional embodiments provide modified variants of SEQ ID NO: 1 to SEQ ID NO: 2, as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine methylation patterns within SEQ ID NO: 1 to SEQ ID NO: 2.

An objective of the invention comprises analysis of the methylation state of one or more CpG dinucleotides within at least one sequence selected form the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2 and sequences complementary thereto.

The disclosed invention provides treated nucleic acids, derived from genomic SEQ ID NO: 1 **to** SEQ ID NO: 2, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization. The genomic sequences in question may comprise one, or more consecutive methylated CpG positions. Said treatment preferably comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof. In a preferred embodiment of the invention, the invention provides a non-naturally occurring modified nucleic acid comprising a sequence of at least 16 contiguous nucleotide bases in length of a sequence selected from the group consisting of SEQ ID NO: 3 to SEQ ID NO: 10. In further preferred embodiments of the invention said nucleic acid is at least 50, 100, 150, 200, 250 or 500 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 3 to SEQ ID NO: 10. Particularly preferred is a nucleic acid molecule that is not identical or complementary to all or a portion of the sequences SEQ ID NO: 1 to SEQ ID NO: 2 or other naturally occurring DNA.

It is preferred that said sequence comprises at least one CpG, TpA or CpA dinucleotide and sequences complementary thereto. The sequences of SEQ ID NO: 3 to SEQ ID NO: 10 provide non-naturally occurring modified versions of the nucleic acid according to SEQ ID NO: 1 to SEQ ID NO: 2, wherein the modification of each genomic sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from said genomic sequence as follows. For each sense strand genomic DNA, *e.g.*, SEQ ID NO:1, four converted versions are disclosed. A first version wherein "C" is converted to "T," but "CpG" remains "CpG" (*i.e.*, corresponds to case where, for the genomic sequence, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); a second version discloses the complement of the disclosed genomic DNA sequence (i.e. *anti*sense strand), wherein "C" is converted to "T," but "CpG" remains "CpG" (*i.e.*, corresponds to case where, for all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted). The 'upmethylated' converted sequences of SEQ ID NO: 1 to SEQ ID NO: 2 correspond to SEQ ID NO: 3 to 6. A third chemically converted version of each genomic sequences is provided, wherein "C" is converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (*i.e.*, corresponds to case where, for the genomic sequences, all "C" residues of CpG dinucleotide sequences are *un*methylated); a final chemically converted version of each sequence, discloses the complement of the disclosed genomic DNA sequence (i.e. *anti*sense strand), wherein "C" is converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (*i.e.*, corresponds to case where, for the complement (*anti*sense strand) of each genomic sequence, all "C" residues of CpG dinucleotide sequences are *un*methylated). The 'downmethylated' converted sequences of SEQ ID NO: 1 to SEQ ID NO: 2 correspond to SEQ ID NO: 7 to 10.

Significantly, *heretofore*, the nucleic acid sequences and molecules according SEQ ID NO: 3 to SEQ ID NO: 10 were not implicated in or connected with the detection, classification or treatment of cellular proliferative disorders.

In an alternative preferred embodiment, the invention further provides oligonucleotides or oligomers suitable for use in the methods of the invention for detecting the cytosine methylation state within genomic or treated (chemically modified) DNA, according to SEQ ID NO: 1 to SEQ ID NO: 2, SEQ ID NO: 3 to SEQ ID NO: 10. Said oligonucleotide or oligomer nucleic acids provide novel diagnostic means. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which is identical to, hybridizes, under moderately stringent or stringent conditions (as defined herein above), to a treated nucleic acid sequence according to SEQ ID NO: 3 to SEQ ID NO: 10 and/or sequences complementary thereto, or to a genomic sequence according to SEQ ID NO: 1 to SEQ ID NO: 2, and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (*e.g.*, oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridize under moderately stringent and/or stringent hybridization conditions to all or a portion of a sequence selected form the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2, SEQ ID NO: 3 to SEQ ID NO: 10 or to the complements thereof. Particularly preferred is a nucleic acid molecule that hybridizes under moderately stringent and/or stringent hybridization conditions to all or a portion of the sequences SEQ ID NO: 3 to SEQ ID NO: 10 but not SEQ ID NO: 1 to SEQ ID NO: 2 or other human genomic DNA.

The identical or hybridizing portion of the hybridizing nucleic acids is typically at least 9, 16, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

Preferably, the hybridizing portion of the inventive hybridizing nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of a sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2, SEQ ID NO: 3 to SEQ ID NO: 10, or to the complements thereof.

Hybridizing nucleic acids of the type described herein can be used, for example, as a primer (*e.g.*, a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridization of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or Tm, which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO: 1 to SEQ ID NO: 2 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridization occurs with a particular concentration of salt (*e.g.*, SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C decrease in the Tm, the temperature of the final wash in the hybridization reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash temperature is decreased by 5°C). In practice, the change in Tm can be between 0.5°C and 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, *e.g.*, SEQ ID NO:1, include those corresponding to sets (sense and antisense sets) of consecutively overlapping oligonucleotides of length X, where the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:
n to (n + (X-1));
where n=1, 2, 3,...(Y-(X-1));
where Y equals the length (nucleotides or base pairs) of SEQ ID NO: 1 (52626);
where X equals the common length (in nucleotides) of each oligonucleotide in the set (*e.g.*, X=20 for a set of consecutively overlapping 20-mers); and
where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to Y- (X-1). For example Z= 52626-19 = 52607 for either sense or antisense sets of SEQ ID NO:1, where X=20.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide. Examples of inventive 20-mer oligonucleotides include the following set of oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO:1:
1-20, 2-21, 3-22, 4-23, 5-24, ... and 52607-52626.
Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

Likewise, examples of inventive 25-mer oligonucleotides include the following set of 219885 oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO: 1:
1-25, 2-26, 3-27, 4-28, 5-29, ... and 52602- 52626.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

The present invention encompasses, for *each* of SEQ ID NO: 3 to SEQ ID NO: 10, SEQ ID NO: 1 to SEQ ID NO: 2 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X, where, *e.g.*, X= 9, 10, 17, 20, 22, 23, 25, 27, 30 or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequences selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 2. Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO: 1 to SEQ ID NO: 2, SEQ ID NO: 3 to SEQ ID NO: 10 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinculeotide) sequences is within the middle third of the oligonucleotide; that is, where the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophors, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridization-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a chromophore, fluorophor, peptide, hybridization-triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligonucleotides or oligomers according to particular embodiments of the present invention are typically used in 'sets,' which contain at least one oligomer for analysis of each of the CpG dinucleotides of a genomic sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2 and sequences complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the treated nucleic acids according to SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto. However, it is anticipated that for economic or other factors it may be preferable to analyse a limited selection of the CpG dinucleotides within said sequences, and the content of the set of oligonucleotides is altered accordingly.

Therefore, in particular embodiments, the present invention provides a set of at least two (2) (oligonucleotides and/or PNA-oligomers) useful for detecting the cytosine methylation state in treated genomic DNA (SEQ ID NO: 3 to SEQ ID NO: 10), or in genomic DNA (SEQ ID NO: 1 to SEQ ID NO: 2 and sequences complementary thereto). These probes enable the differentiation of pre-cancerous (i.e. malignant or pre-malignant) colorectal lesions from benign colorectal lesions (commonly referred to as benign). The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in treated genomic DNA (SEQ ID NO: 3 to SEQ ID NO: 10), or in genomic DNA (SEQ ID NO: 1 to SEQ ID NO: 2 and sequences complementary thereto).

In preferred embodiments, at least one, and more preferably all members of a set of oligonucleotides is bound to a solid phase.

In further embodiments, the present invention provides a set of at least two (2) oligonucleotides that are used as 'primer' oligonucleotides for amplifying DNA sequences of one of SEQ ID NO: 1 to SEQ ID NO: 2, SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto, or segments thereof.

It is anticipated that the oligonucleotides may constitute all or part of an "array" or "DNA chip" (*i.e.*, an arrangement of different oligonucleotides and/or PNA-oligomers bound to a solid phase). Such an array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized, for example, in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid-phase surface may be composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold. Nitrocellulose as well as plastics such as nylon, which can exist in the form of pellets or also as resin matrices, may also be used. An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of Nature Genetics (*Nature Genetics Supplement*, Volume 21, January 1999, and from the literature cited therein). Fluorescently labelled probes are often used for the scanning of immobilized DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly suitable for fluorescence labels. The

detection of the fluorescence of the hybridised probes may be carried out, for example, via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

It is also anticipated that the oligonucleotides, or particular sequences thereof, may constitute all or part of an "virtual array" wherein the oligonucleotides, or particular sequences thereof, are used, for example, as 'specifiers' as part of, or in combination with a diverse population of unique labeled probes to analyze a complex mixture of analytes. Such a method, for example is described in US 2003/0013091 (United States serial number 09/898,743, published 16 January 2003). In such methods, enough labels are generated so that each nucleic acid in the complex mixture (*i.e.*, each analyte) can be uniquely bound by a unique label and thus detected (each label is directly counted, resulting in a digital read-out of each molecular species in the mixture).

It is particularly preferred that the oligomers according to the invention are utilized for the early detection of colorectal cancer by differentiating benign colorectal lesions from those undergoing malignant transformation.

In the most preferred embodiment of the method, the presence or absence of a colon lesion undergoing malignant transformation is determined and/or the differentiation between malignant/pre-malignant and benign lesions is made. This is achieved by analysis of the methylation status of at least one target sequence comprising at least one CpG position said sequence comprising, or hybridizing under stringent conditions to at least 16 contiguous nucleotides of a sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2 and complements thereof. The present invention further provides a method for ascertaining genetic and/or epigenetic parameters of the genomic sequence according to SEQ ID NO: 1 to SEQ ID NO: 2 within a subject by analyzing cytosine methylation and single nucleotide polymorphisms. Said method comprising contacting a nucleic acid comprising SEQ ID NO: 1 to SEQ ID NO: 2 in a biological sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

In a preferred embodiment, said method comprises the following steps: In the *first step*, a sample of the tissue to be analyzed is obtained. The source may be any suitable source, such as cell lines, histological slides, biopsies, paraffin-embedded tissue, body fluids, stool, colonic effluent, urine, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood and all possible combinations thereof. It is preferred that said sources of DNA are stool or body fluids selected from the group consisting colonic effluent, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood.

The genomic DNA is then isolated from the sample. Genomic DNA may be isolated by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants e.g. by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense and required quantity of DNA.

Wherein the sample DNA is not enclosed in a membrane (e.g. circulating DNA from a blood sample) methods standard in the art for the isolation and/or purification of DNA may be employed. Such methods include the use of a protein degenerating reagent *e.g.*, chaotropic salt e.g. guanidine hydrochloride or urea; or a detergent e.g. sodium dodecyl sulphate (SDS), cyanogen bromide. Alternative methods include but are not limited to ethanol precipitation or propanol precipitation, vacuum concentration amongst others by means of a centrifuge. The person skilled in the art may also make use of devices such as filter devices, *e.g.*, ultrafiltration, silica surfaces or membranes, magnetic particles, polystyrene particles, polystyrene surfaces, positively charged surfaces, and positively charged membranes, charged membranes, charged surfaces, charged switch membranes, charged switched surfaces.

Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

In the *second step* of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be understood as 'pre-treatment' or 'treatment' hereinafter.

This is preferably achieved by means of treatment with a bisulfite reagent. The term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences. Methods of said treatment are known in the art (e.g., PCT/EP2004/011715, which is incorporated by reference in its entirety). It is preferred that the bisulfite treatment is conducted in the presence of denaturing solvents such as but not limited to n-alkylenglycol, particularly diethylene glycol dimethyl ether (DME), or in the presence of dioxane or dioxane derivatives. In a preferred embodiment the denaturing solvents are used in concentrations between 1% and 35% (v/v). It is also preferred that 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 trihydroxybenzoe acid

and derivates 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 periods of times during the reaction (see: PCT/EP2004/011715 which is incorporated by reference in its entirety). The bisulfite treated DNA is preferably purified priori to the quantification. This may be conducted by any means known in the art, such as but not limited to ultrafiltration, preferably carried out by means of Microcon^(TM) columns (manufactured by Millipore^(TM)). The purification is carried out according to a modified manufacturer's protocol (see: PCT/EP2004/011715 which is incorporated by reference in its entirety).

In the *third step* of the method, fragments of the treated DNA are amplified, using sets of primer oligonucleotides according to the present invention, 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). Preferably said amplificates are 100 to 2,000 base pairs in length. The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridise under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one of SEQ ID NO: 3 TO SEQ ID NO: 10 and sequences complementary thereto.

In an alternate embodiment of the method, the methylation status of pre-selected CpG positions within at least one nucleic acid sequences selected from the group consisting SEQ ID NO: 1 TO SEQ ID NO: 2 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent 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 primers pairs contain at least one primer which 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. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 3 TO SEQ ID NO: 10 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide .A further preferred embodiment of the method comprises the use of *blocker* oligonucleotides (the HeavyMethyl™ assay). The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. Blocking probe oligonucleotides are hybridised to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The

probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'CpA' or 'TpA' at the position in question, as opposed to a 'CpG' if the suppression of amplification of methylated nucleic acids is desired.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivitized at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-terminii thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (*e.g.*, with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker—a process that normally results in degradation of the hybridized blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

Preferably, therefore, the base sequence of said *blocking oligonucleotides* is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto, wherein the base sequence of said oligonucleotides comprises at least one CpG, TpG or CpA dinucleotide.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplificates have a single positive or negative net charge,

allowing for better delectability in the mass spectrometer. The detection may be carried out and visualized by means of, *e.g.*, matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas and Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapor phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut and Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionization process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallization. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut and Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In the *fourth step* of the method, the amplificates obtained during the third step of the method are analyzed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplificates were obtained by means of MSP amplification, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primer, according to the base sequences of said primer.

Amplificates obtained by means of both standard and methylation specific PCR may be further analyzed by means of based-based methods such as, but not limited to, array technology and probe

based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplificates synthesized in *step three* are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplificates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the present Sequence Listing; and the segment comprises at least one CpG , TpG or CpA dinucleotide. The hybridizing portion of the hybridizing nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within a sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2, and the equivalent positions within SEQ ID NO: 3 to SEQ ID NO: 10.

Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridized amplificates are then removed. The hybridized amplificates are then detected. In this context, it is preferred that labels attached to the amplificates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes (as detailed above) that are hybridized to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; *also see* United States Patent No. 6,331,393) employing a dual-labeled fluorescent oligonucleotide probe (TaqMan™ PCR, using an ABI Prism 7700 Sequence Detection System, Perkin Elmer Applied Biosystems, Foster City, California). The TaqMan™ PCR reaction employs the use of a non-extendible interrogating oligonucleotide, called a TaqMan™ probe, which, in preferred embodiments, is designed to hybridize

to a CpG-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (*e.g.*, phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLightTM™ assay. Variations on the TaqMan™ detection methodology that are also suitable for use with the described invention include the use of dual-probe technology (Lightcycler™) or fluorescent amplification primers (Sunrise™ technology). Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

In a further preferred embodiment of the method, the *fourth step* of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

In yet a further embodiment of the method, the *fourth step* of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the *third step* of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

Best mode

In the most preferred embodiment of the method the genomic nucleic acids are isolated and treated according to the first three steps of the method outlined above, namely:

a) obtaining, from a subject, a biological sample having subject genomic DNA;

b) extracting or otherwise isolating the genomic DNA;

c) treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties; and wherein

d) amplifying subsequent to treatment in c) is carried out in a methylation specific manner, namely by use of methylation specific primers or *blocking oligonucleotides*, and further wherein

e) detecting of the amplificates is carried out by means of a real-time detection probe, as described above.

Preferably, where the subsequent amplification of d) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto, wherein the base sequence of said

oligomers comprise at least one CpG dinucleotide.

Step e) of the method, namely the detection of the specific amplificates indicative of the methylation status of one or more CpG positions of at least one sequences of the group comprising SEQ ID NO: 1 to SEQ ID NO: 2 is carried out by means of *real-time* detection methods as described above.

Additional embodiments of the invention provide a method for the analysis of the methylation status of genomic DNA according to the invention (SEQ ID NO: 1 to SEQ ID NO: 2, and complements thereof) without the need for bisulfite conversion. Methods are known in the art wherein a methylation sensitive restriction enzyme reagent, or a series of restriction enzyme reagents comprising methylation sensitive restriction enzyme reagents that distinguishes between methylated and non-methylated CpG dinucleotides within a target region are utilized in determining methylation, for example but not limited to DMH.

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

Once the nucleic acids have been extracted, the genomic double-stranded DNA is used in the analysis.

In a preferred embodiment, the DNA may be cleaved prior to treatment with methylation sensitive restriction enzymes. Such methods are known in the art and may include both physical and enzymatic means. Particularly preferred is the use of one or a plurality of restriction enzymes which are not methylation sensitive, and whose recognition sites are AT rich and do not comprise CG dinucleotides. The use of such enzymes enables the conservation of CpG islands and CpG rich regions in the fragmented DNA. The non-methylation-specific restriction enzymes are preferably selected from the

group consisting of MseI, BfaI, Csp6I, TruII, TvuII, Tru9I, Tvu9I, MaeI and XspI. Particularly preferred is the use of two or three such enzymes. Particularly preferred is the use of a combination of MseI, BfaI and Csp6I.

The fragmented DNA may then be ligated to adaptor oligonucleotides in order to facilitate subsequent enzymatic amplification. The ligation of oligonucleotides to blunt and sticky ended DNA fragments is known in the art, and is carried out by means of dephosphorylation of the ends (e.g. using calf or shrimp alkaline phosphatase) and subsequent ligation using ligase enzymes (e.g. T4 DNA ligase) in the presence of dATPs. The adaptor oligonucleotides are typically at least 18 base pairs in length.

In the *third step*, the DNA (or fragments thereof) is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4. Preferably the methylation status of both Septin 9 (including any transcript variants thereof) and ALX4 are analyzed.

Preferably, the methylation-specific restriction enzyme is selected from the group consisting of *Bsi E1, Hga I HinPl, Hpy99I, Ava I, Bce AI, Bsa HI, BisI, BstUI, BshI236I, AccII, BstFNI, McrBC, GlaI, MvnI, HpaII (HapII), HhaI, AciI, SmaI, HinP1I, HpyCH4IV, EagI* and mixtures of two or more of the above enzymes. Preferred is a mixture containing the restriction enzymes BstUI, HpaII, HpyCH4IV and HinP1I.

In the *fourth step*, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and said amplificates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels. Particularly preferred is amplification by means of an amplification enzyme and at least two primers comprising, in each case a contiguous sequence at least 16 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting SEQ ID NO: 1 TO SEQ ID NO: 2, and complements thereof. Preferably said contiguous sequence is at least 16, 20 or 25 nucleotides in length. In an alternative embodiment said primers may be complementary to any adaptors linked to the fragments.

In the *fifth step* the amplificates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridization analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis. Preferably said detection is carried out by hybridization to at least one nucleic acid or peptide nucleic acid comprising

in each case a contiguous sequence at least 16 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2, and complements thereof. Preferably said contiguous sequence is at least 16, 20 or 25 nucleotides in length.

Subsequent to the determination of the methylation state or level of the genomic nucleic acids the class of cellular proliferative disorder (benign or malignant) is deduced based upon the methylation state or level of at least one CpG dinucleotide sequence of at least one sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of at least one sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2 wherein methylation is associated with a colorectal lesion undergoing malignant transformation. Wherein said methylation is determined by quantitative means the cut-off point for determining said the presence of methylation is preferably zero (i.e. wherein a sample displays any degree of methylation it is determined as having a methylated status at the analyzed CpG position). Nonetheless, it is foreseen that the person skilled in the art may wish to adjust said cut-off value in order to provide an assay of a particularly preferred sensitivity or specificity. Accordingly said cut-off value may be increased (thus increasing the specificity), said cut off value may be within a range selected form the group consisting of 0%-5%, 5%-10%, 10%-15%, 15%-20%, 20%-30% and 30%-50%. Particularly preferred are the cut-offs 10%, 15%, 25%, and 30%.

Prognostic Assays for cellular proliferative disorders

The present invention enables diagnosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and/or ALX4 may be used as markers. Said parameters obtained by means of the present invention may be compared to another set of genetic and/or epigenetic parameters, the differences serving as the basis for a diagnosis and/or prognosis of events which are disadvantageous to patients or individuals.

More specifically the present invention enables the screening of at-risk populations for the early detection of cancers, most preferably colorectal carcinomas. Furthermore, the present invention enables the differentiation of malignant or pre-malignant lesions from those which are likely to remain benign (i.e. non-cancerous).

Specifically, the present invention provides for colorectal neoplasm detection assays based on measurement of differential expression (preferably methylation) of one or more CpG dinucleotide sequences of at least one sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2 that comprise such a CpG dinucleotide sequence. Typically, such assays involve obtaining a sample

from a subject, performing an assay to measure the expression of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4, preferably by determining the methylation status of at least one sequence selected from the group consisting SEQ ID NO: 1 to SEQ ID NO: 2, derived from the sample, relative to a control sample, or a known standard and making a diagnosis based thereon. It is particularly preferred that the methylation status of both Septin 9 (including any transcript variants thereof) and ALX4 are analyzed.

In particular preferred embodiments, inventive oligomers are used to assess the CpG dinucleotide methylation status, such as those based on SEQ ID NO: 1 to SEQ ID NO: 2, SEQ ID NO: 3 to SEQ ID NO: 10, or arrays thereof, as well as in kits based thereon and useful for the classification of colorectal cell proliferative disorders.

Kits

Moreover, an additional aspect of the present invention is a kit comprising: a means for determining methylation of at least one gene selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4. The means for determining methylation comprise preferably a bisulfite-containing reagent; one or a plurality of oligonucleotides consisting whose sequences in each case are identical, are complementary, or hybridise under stringent or highly stringent conditions to a 9 or more preferably up to 18 base long segment of a sequence selected from SEQ ID NO: 3 to SEQ ID NO: 10; and optionally instructions for carrying out and evaluating the described method of methylation analysis. In one embodiment the base sequence of said oligonucleotides comprises at least one CpG, CpA or TpG dinucleotide.

In a further embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLight™, HeavyMethyl, COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

In a preferred embodiment the kit may comprise additional bisulfite conversion reagents selected from the group consisting: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

In a further alternative embodiment, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimised for primer extension mediated by the polymerase, such as PCR. In another embodiment of the invention the kit further comprising means for obtaining a

biological sample of the patient. Preferred is a kit, which further comprises a container suitable for containing the means for determining methylation of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4 in the biological sample of the patient, and most preferably further comprises instructions for use and interpretation of the kit results. In a preferred embodiment the kit comprises: (a) a bisulfite reagent; (b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides containing two oligonucleotides whose sequences in each case are identical, are complementary, or hybridise under stringent or highly stringent conditions to a 9 or more preferably up to 18 base long segment of a sequence selected from SEQ ID NO: 3 to SEQ ID NO: 10; and optionally (d) instructions for use and interpretation of the kit results. In an alternative preferred embodiment the kit comprises: (a) a bisulfite reagent; (b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient; (c) at least one oligonucleotides and/or PNA-oligomer having a length of at least 9 or 16 nucleotides which is identical to or hybridises to a pre-treated nucleic acid sequence according to one of SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto; and optionally (d) instructions for use and interpretation of the kit results.

In an alternative embodiment the kit comprises: (a) a bisulfite reagent; (b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides containing two oligonucleotides whose sequences in each case are identical, are complementary, or hybridise under stringent or highly stringent conditions to a 9 or more preferably 18 base long segment of a sequence selected from SEQ ID NO: 3 to SEQ ID NO: 10; (d) at least one oligonucleotides and/or PNA-oligomer having a length of at least 9 or 16 nucleotides which is identical to or hybridises to a pre-treated nucleic acid sequence according to one of SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto; and optionally (e) instructions for use and interpretation of the kit results.

The kit may also contain other components such as buffers or solutions suitable for blocking, washing or coating, packaged in a separate container.

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 at least one gene selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4; restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and labeled nucleotides. Typical reagents (*e.g.*, as might be found in a typical MethyLight™-based kit) for MethyLight™ analysis may include, but are not limited to: PCR primers for the bisulfite converted sequence of at least one gene or genomic sequence selected from the group consisting of

Septin 9 (including all transcript variants thereof) and ALX4; bisulfite specific probes (*e.g.*, TaqMan ™ or Lightcycler ™); optimized PCR buffers and deoxynucleotides; and Taq polymerase.

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 gene (or bisulfite treated DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE™ primers for the bisulfite converted sequence of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4; reaction buffer (for the Ms-SNuPE reaction); and labeled nucleotides.

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 the bisulfite converted sequence of or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4, optimized PCR buffers and deoxynucleotides, and specific probes.

Moreover, an additional aspect of the present invention is an alternative kit comprising a means for determining methylation of at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and ALX4, wherein said means comprise preferably at least one methylation specific restriction enzyme; one or a plurality of primer oligonucleotides (preferably one or a plurality of primer pairs) suitable for the amplification of a sequence comprising at least one CpG dinucleotide of a sequence selected from SEQ ID NO: 1 to SEQ ID NO: 2; and optionally instructions for carrying out and evaluating the described method of methylation analysis. In one embodiment the base sequence of said oligonucleotides are identical, are complementary, or hybridise under stringent or highly stringent conditions to an at least 18 base long segment of a sequence selected from SEQ ID NO: 1 to SEQ ID NO: 2.

In a further embodiment said kit may comprise one or a plurality of oligonucleotide probes for the analysis of the digest fragments, preferably said oligonucleotides are identical, are complementary, or hybridise under stringent or highly stringent conditions to an at least 16 base long segment of a sequence selected from SEQ ID NO: 1 to SEQ ID NO: 2.

In a preferred embodiment the kit may comprise additional reagents selected from the group consisting: buffer (*e.g.*, restriction enzyme, PCR, storage or washing buffers); DNA recovery reagents or kits (*e.g.*, precipitation, ultrafiltration, affinity column) and DNA recovery components.

In a further alternative embodiment, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as

PCR. In another embodiment of the invention the kit further comprising means for obtaining a biological sample of the patient. In a preferred embodiment the kit comprises: (a) a methylation sensitive restriction enzyme reagent; (b) a container suitable for containing the said reagent and the biological sample of the patient; (c) at least one set of oligonucleotides one or a plurality of nucleic acids or peptide nucleic acids which are identical, are complementary, or hybridise under stringent or highly stringent conditions to an at least 9 base long segment of a sequence selected from SEQ ID NO: 1 to SEQ ID NO: 2; and optionally (d) instructions for use and interpretation of the kit results.

In an alternative preferred embodiment the kit comprises: (a) a methylation sensitive restriction enzyme reagent; (b) a container suitable for containing the said reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides suitable for the amplification of a sequence comprising at least one CpG dinucleotide of a sequence selected from SEQ ID NO: 1 to SEQ ID NO: 2; and optionally (d) instructions for use and interpretation of the kit results.

In an alternative embodiment the kit comprises: (a) a methylation sensitive restriction enzyme reagent; (b) a container suitable for containing the said reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides suitable for the amplification of a sequence comprising at least one CpG dinucleotide of a sequence selected from SEQ ID NO: 1 to SEQ ID NO: 2; (d) at least one set of oligonucleotides one or a plurality of nucleic acids or peptide nucleic acids which are identical , are complementary, or hybridize under stringent or highly stringent conditions to an at least 9 base long segment of a sequence selected from SEQ ID NO: 1 to SEQ ID NO: 2 and optionally (e) instructions for use and interpretation of the kit results.

The kit may also contain other components such as buffers or solutions suitable for blocking, washing or coating, packaged in a separate container.

The invention further relates to a kit for use in detecting and/or providing a classification of colorectal lesions in a subject by means of methylation-sensitive restriction enzyme analysis. Said kit comprises a container and a DNA microarray component. Said DNA microarray component being a surface upon which a plurality of oligonucleotides are immobilized at designated positions and wherein the oligonucleotide comprises at least one CpG methylation site. At least one of said oligonucleotides is specific for the at least one gene or genomic sequence selected from the group consisting of Septin 9 (including all transcript variants thereof) and/or ALX4 and comprises a sequence of at least 15 base pairs in length but no more than 200 bp of a sequence according to one of SEQ ID NO: 1 to SEQ ID NO: 2. Preferably said sequence is at least 15 base pairs in length but no more than 80 bp of a sequence according to one of SEQ ID NO: 1 to SEQ ID NO: 2. It is further preferred that said sequence is at least 20 base pairs in length but no more than 30 bp of a sequence according to one of

SEQ ID NO: 1 to SEQ ID NO: 2.

Said test kit preferably further comprises a restriction enzyme component comprising one or a plurality of methylation-sensitive restriction enzymes.

In a further embodiment, said test kit is further characterized in that it comprises at least one methylation-specific restriction enzyme, and wherein the oligonucleotides comprise a restriction site of said at least one methylation specific restriction enzymes.

The kit may further comprise one or several of the following components, which are known in the art for DNA enrichment: a protein component, said protein binding selectively to methylated DNA; a triplex-forming nucleic acid component, one or a plurality of linkers, optionally in a suitable solution; substances or solutions for performing a ligation e.g. ligases, buffers; substances or solutions for performing a column chromatography; substances or solutions for performing an immunology based enrichment (e.g. immunoprecipitation); substances or solutions for performing a nucleic acid amplification e.g. PCR; a dye or several dyes, if applicable with a coupling reagent, if applicable in a solution; substances or solutions for performing a hybridization; and/or substances or solutions for performing a washing step.

The described invention further provides a composition of matter useful for the classification of colorectal lesions. Said composition comprising at least one nucleic acid 18 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 3 to SEQ ID NO: 10, and one or more substances taken from the group comprising : 1-5 mM Magnesium Chloride, 100-500 µM dNTP, 0.5-5 units of taq polymerase, bovine serum albumen, an oligomer in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridizes under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 3 to SEQ ID NO: 10 and sequences complementary thereto. It is preferred that said composition of matter comprises a buffer solution appropriate for the stabilization of said nucleic acid in an aqueous solution and enabling polymerase based reactions within said solution.. Suitable buffers are known in the art and commercially available.

In further preferred embodiments of the invention said at least one nucleic acid is at least 50, 100, 150, 200, 250 or 500 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 3 to SEQ ID NO: 10.

While the present invention has been described with specificity in accordance with certain of its

preferred embodiments, the following examples serve only to illustrate the invention and are not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

Example 1

In the following study, the methylation status of plasma samples derived from patients with varying type and stages of colorectal lesions was analysed in order to confirm lesion specific methylation within the genes Septin 9 and ALX4.

The assays were MSP and HeavyMethyl assays as described above, methylation specific real-time assays for the analysis of bisulfite converted DNA. The assays were designed to be run on the LightCycler platform (Roche Diagnostics), but other such instruments commonly used in the art are also suitable. MSP and HeavyMethyl amplificates were designed to be detected by means of Lightcycler style dual probes. Each assay was run in triplicate on plasma samples obtained from a commercial provider.

Study population

Samples were collected from male and female patients ages 40 to 80 but predominantly ages 50 and older. Case report forms were reviewed by a physician and severity of disease determined. In total, serum was collected from 36 patients with adenomatous polyps (13 female, 23 male, median age 63.5, range 24-75) and 13 patients with hyperplastic polyps (5 female, 8 male, median age 58, range 20-73). Serum was also collected from 22 individuals undergoing colonoscopy for various other reasons without neoplasia or preneoplasia of the colon and rectum (negative control) as well as 5 patients with colorectal carcinoma.

Of the polyp serum samples 7 samples were taken from patients with polyps >1cm with dysplasia, 9 samples from polyps >1cm without dysplasia, 17 samples from polyps <1cm without dysplasia, 3 samples from polyps <1cm with dysplasia, 13 samples from hyperplastic polyps of which 11 were from polyps <1cm and 2 were from polyps >2cm. The colorectal carcinoma samples were obtained from Stage 1 (4 samples) and stage 3 (1 sample) patients.

Plasma methylation analysis

Plasma was processed first by extraction of free-circulating DNA using the Total Nucleic Acid DNA extraction kit (Roche Applied Science) and Roche MagNaPure device. Eluted DNA's from each patient were pooled and concentrated on microcon filters. DNA Methylation information was preserved by deamination of unmethylated cytosines using sodium bisulfite as described above. Bisulfite treated plasma DNA in each sample was quantified on the Roche LC 2.0™ device using a non-methylation specific assay for beta-actin. ALX4 methylation was determined by means of the

MSP assay. Septin 9 methylation was determined using an assay based on the applicant's HM real-time PCR technology (see Cottrell SE, Distler J, et al.. NAR 2004; 32(1):e10). The 90% limit of detection of the *Septin 9* assay was estimated as 21pg by a dilution series of methylated (SSS1 treated) DNA in a background of 50ng blood DNA (Roche human genomic DNA). The Roche LC 2.0 was also used to measure Septin 9 amplification. A plasma equivalent of 1.6 ml to 1.9 ml of DNA was added per PCR reaction and each plasma sample run in duplicate or triplicate.

Assays

Genomic region of interest: SEQ ID NO: 1; Assay type: MSP

Primers: cgtcgcaacgcgtacg (SEQ ID NO: 11); cgcggtttcgattttaatgc (SEQ ID NO: 12)

Lightcycler probes: actccgacttaacccgacgatcg - fluo (SEQ ID NO: 13); LC 640-acgaaattcctaacgcaaccgct-p (SEQ ID NO: 14)

Amplificate sequence:
Cgtcgcaacgcgtacgactcaaaacttaataactccgacttaacccgacgatcgcgacgaaattcctaacgcaaccgcttaaaacttcgcattaaaatcgaaaccgcg (SEQ ID NO: 15)

Temperature cycling program:

activation:	95°C	10min	
55 cycles:	95°C	15 sec	(20°C/s)
	60°C	45 sec	(20°C/s)
	72°C	15 sec	(20°C/s)

Genomic region of interest: SEQ ID NO: 2; Assay type: HeavyMethyl

Primers: GtAGtAGttAGtttAGtAtttAttTT (SEQ ID NO: 16); CCCACCAaCCATCATaT (SEQ ID NO: 17)

Lightcycler probes: Gtt cga aat gat ttt att tag ttg c-FL (SEQ ID NO: 18); LC-Red640-cgt tga tcg cgg ggt tc-PH (SEQ ID NO: 19)

Blocker sequence: CATCATaTCAaACCCCACAaTCAACACACAaC-Inv 3' (SEQ ID NO: 20)

Temperature cycling program:

Activation:	95°C	10min
55 cycles:	95°C	10 sec
	56°C	30 sec
	72°C	10 sec
Cooling:	10°C	30 sec

Results

A replicate was determined as positive if it presented with greater than 4% methylation, a sample was determined as positive according to how many of the replicates presented methylation.

Single gene analysis

Septin 9					
		2/3 positive		1/3 positive	
	Total sample	#	%	#	%
Normal	22	1	4.5	4	18
Polyp	49	6	12	17	35
CRC	5	2	40	2	40

ALX4					
		2/3 positive		1/3 positive	
	Total sample	#	%	#	%
Normal	22	4	18	14	64
Polyp	49	23	47	38	77.5
CRC	5	2	40	4	80

Panel analysis (Septin 9 + ALX4)

		2/3 positive		3/6 positive	
	Total # samples	#	%	#	%
Normal	22	4	18	2	9
Polyp	49	26	53	24	49
CRC	5	3	60	3	60

According to polyp histology

		Septin 9 2/3		ALX 4 2/3		Septin 9 + ALX4 2/3		Septin 9 + ALX4 3/6	
Polyp Characteristics	# samples	#	%	#	%	#	%	#	%
>or = 1cm	18	3	17	10	55	13	72	12	67
< 1cm	31	3	10	13	42	13	42	12	39
Tubular or villous adenoma	36	5	14	17	47	20	55	20	55
Tubular or villous adenoma +>1cm	11	3	27	8	73	11	100	11	100
Hyperplastic polyps	13	1	8	6	46	6	46	4	31

Furthermore, of the 10 adenomas with intraepithelial lesions 80% displayed methylation in at least 3 out of 6 replicates of Septin 9 and ALX4.

Conclusions

It was determined that the sensitivity of a real-time PCR assay for detection of polyps larger than 1cm was 23%. Our results indicate that the *Septin 9* biomarker is also highly specific (95%) in asymptomatic individuals over 50 years of age.

The results of combining Septin 9 and ALX4 in a marker panel indicate that sensitivity for detecting polyps can be considerably improved while maintaining high specificity. The marker panel detected polyps larger than 1cm with a sensitivity of 67%. Sensitivity for large adenomas (> 1cm) or adenomas with IEN showed even greater improvement (100%, 80% respectively). Specificity of the panel in asymptomatic individuals over 50 years of age was 91%.Patient compliance and performance of current screening strategies limit the effectiveness of tests available on the market today. An easily administered blood-based test for early detection of colorectal cancer followed by colonoscopy for positive individuals has the potential to be a very effective tool for reducing mortality from this disease.

Table 1

Genomic SEQ ID NO:	Ensembl database genomic location	Associated gene transcript(s)*	Methylated bisulfite converted sequence (sense)	Methylated bisulfite converted sequence (antisense)	Unmethylated bisulfite converted sequence (sense)	Unmethylated bisulfite converted sequence (antisense)
1	Chromosome 11:44238570:44291195:1	ALX4	3	4	7	8
2	Chromosome 17:72786362:73008270:1	Septin 9	5	6	9	10

Claims

1. A method for detecting and/or classifying pre-cancerous colorectal cell proliferative disorders in a subject comprising determining the presence of CpG methylation of Septin 9, wherein CpG methylation is indicative of the presence of pre-cancerous colorectal cell proliferative disorders.

2. A method for detecting and/or classifying pre-cancerous colorectal cell proliferative disorders in a subject, comprising contacting genomic DNA isolated from a biological sample obtained from said subject with at least one reagent, or series of reagents that distinguishes between methylated and nonmethylated CpG dinucleotides within at least one target region of the genomic DNA, wherein the target region comprises, or hybridizes under stringent conditions to a sequence of at least 16 contiguous nucleotides of SEQ ID NO: 2, wherein said contiguous nucleotides comprise at least one CpG dinucleotide sequence, and whereby detecting and/or classifying pre-cancerous colorectal cell proliferative disorders is, at least in part, afforded.

3. A method for detecting and/or classifying pre-cancerous colorectal cell proliferative disorders, comprising:

   a) extracting or otherwise isolating genomic DNA from a biological sample obtained from the subject;
   b) treating the genomic DNA of a), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;
   c) contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least one primer comprising, a contiguous sequence of at least 9 nucleotides that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 9, SEQ ID NO: 10, and complements thereof, wherein the treated genomic DNA or the fragment thereof is either amplified to produce at least one amplificate, or is not amplified; and
   d) determining, based on a presence or absence of, or on a property of said amplificate, the methylation state or level of at least one CpG dinucleotide of SEQ ID NO: 2, or an average, or a value reflecting an average methylation state or level of a plurality of CpG dinucleotides of SEQ ID NO: 2, whereby at

least one of detecting and classifying pre-cancerous colorectal cellular proliferative disorders is, at least in part, afforded.

4. The method of claim 3, wherein treating the genomic DNA, or the fragment thereof in b), comprises use of a reagent selected from the group comprising of bisulfite, hydrogen sulfite, disulfite, and combinations thereof.

5. The method of claim 4, wherein contacting or amplifying in c) comprises use of at least one method selected from the group comprising: use of a heat-resistant DNA polymerase as the amplification enzyme; use of a polymerase lacking 5'-3' exonuclease activity; use of a polymerase chain reaction (PCR); generation of an amplificate nucleic acid molecule carrying a detectable label.

6. The method of any of claim 1 to 5, wherein the biological sample obtained from the subject is selected from the group comprising cell lines, histological slides, biopsies, paraffin-embedded tissue, body fluids, stool, colonic effluent, urine, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood and combinations thereof.

7. The method of claim 6, further comprising in step d) the use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 9, SEQ ID NO: 10, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized.

8. The method of claim 5, wherein determining in d) comprises hybridization of at least one nucleic acid molecule or peptide nucleic acid molecule in each case comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 9, SEQ ID NO: 10, and complements thereof.

9. The method of claim 8, wherein at least one such hybridizing nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase.

10. The method of claim 8, further comprising extending at least one such hybridized nucleic acid molecule by at least one nucleotide base.

11. The method of claim 5, wherein determining in d), comprises sequencing of the amplificate.

12. The method of claim 5, wherein contacting or amplifying in c), comprises use of methylation specific primers.

13. A method for detecting and/or classifying pre-cancerous colorectal cellular proliferative disorders, comprising:

a) extracting or otherwise isolating genomic DNA from a biological sample obtained from the subject,

b) digesting the genomic DNA of a), or a fragment thereof, with one or more methylation sensitive restriction enzymes, contacting the DNA restriction enzyme digest of b), with an amplification enzyme and at least two primers suitable for the amplification of a sequence comprising at least one CpG dinucleotide of SEQ ID NO: 2, and

c) determining, based on a presence or absence of an amplificate the methylation state or level of at least one CpG dinucleotide of SEQ ID NO: 2, whereby at least one of detecting and classifying pre-cancerous cellular proliferative disorders is, at least in part, afforded.

14. The method according to claim 13 wherein the presence or absence of an amplificate is determined by means of hybridization to at least one nucleic acid or peptide nucleic acid which is identical, complementary, or hybridizes under stringent or highly stringent conditions to an at least 16 base long segment of SEQ ID NO: 2.

15. A treated nucleic acid derived from genomic SEQ ID NO: 2, when used for the detection of pre-cancerous colorectal cell proliferative disorders, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization.

16. An isolated nucleic acid, when used for the detection of pre-cancerous colorectal cell proliferative disorders, comprising at least 16 contiguous nucleotides of a treated genomic DNA sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 9 and SEQ ID NO: 10, and sequences complementary thereto, wherein the treatment is suitable to convert at least one unmethylated cytosine base of

the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization.

17. An isolated nucleic acid, when used for the detection of pre-cancerous colorectal cell proliferative disorders, comprising at least 50 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 9 and SEQ ID NO: 10, and sequences complementary thereto.

18. A nucleic acid, when used for the detection of pre-cancerous colorectal cell proliferative disorders, comprising at least 16 contiguous nucleotides of a treated genomic DNA sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 9 and SEQ ID NO: 10 and sequences complementary thereto as a diagnostic means.

19. The nucleic acid of any of claims 15 to 18 wherein the contiguous base sequence comprises at least one CpG, TpG or CpA dinucleotide sequence.

20. A kit when used in the method according to claim 2, comprising

a) a bisulfite reagent;
b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient;
c) at least one set of oligonucleotides containing two oligonucleotides whose sequences in each case are identical, are complementary, or hybridize under stringent or highly stringent conditions to a 9 or more preferably 18 base long segment of a sequence selected from SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 9 and SEQ ID NO: 10.

21. A kit when used in the method according to claim 2, comprising

a) a methylation sensitive restriction enzyme reagent;
b) a container suitable for containing the said reagent and the biological sample of the patient;
c) at least one set of oligonucleotides one or a plurality of nucleic acids or peptide nucleic acids which are identical, are complementary, or hybridize under stringent or highly stringent conditions to an at least 9 base long segment of SEQ 10 NO: 2; and optionally
d) instructions for use and interpretation of the kit results.

The Sequence Listing for this application has been submitted to WIPO in electronic format and can be obtained upon request from the International Bureau or found on the WIPO website at http://www.wipo.int/pctdb/en/sequences/