

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 July 2009 (09.07.2009)

PCT

(10) International Publication Number
WO 2009/084960 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/574* (2006.01)

(74) Agent: HATZMANN, M., J.; Vereenigde, Johan de Witt-
laan 7, NL-2517 JR Den Haag (NL).

(21) International Application Number:
PCT/NL2008/050857

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, 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, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date:
29 December 2008 (29.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
07150458.3 28 December 2007 (28.12.2007) EP

(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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*):
ACADEMISCH ZIEKENHUIS LEIDEN [NL/NL];
Albinusdreef 2, NL-2333 ZA Leiden (NL).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **BOER, Judith Mary** [NL/NL]; H. Heijermansplein 19, NL-2024 JK Haarlem (NL). **VAN ROON, Eddy Herman Jasper** [NL/NL]; Donizettihof 5, NL-2402 EH Alphen Aan Den Rijn (NL). **MORREAU, Johannes** [NL/NL]; Koepellaan 4a, NL-2061 CV Bloemendaal (NL).

Published:

— *without international search report and to be republished upon receipt of that report*

(54) Title: METHYLATION DETECTION IN THE GENOMIC REGION OF A RECEPTOR PROTEINTYROSINE PHOSPHATASE GAMMA GENE FOR DETECTION AND/OR DIAGNOSIS OF A TUMOUR

(57) Abstract: The invention discloses means and methods for detecting methylation of a CpG in a region of a receptor protein tyrosine phosphatase gamma gene in samples obtained from a relevant site of an individual. The means and methods can be used to determine whether said individual is suffering from a tumour.



WO 2009/084960 A2

P78692PC00

Title: Methylation detection in the genomic region of a receptor protein-tyrosine phosphatase gamma gene for detection and/or diagnosis of a tumour.

5 The invention relates to the field of medicine and diagnosis. The invention in particular relates to methylation of genomic DNA and the correlation with the presence of cancer cells or precursors thereof in a sample. The invention further relates to the use of the genomic region of receptor protein-tyrosine phosphatase gamma gene therein and the use of
10 said genomic region to screen for CpG methylation involved in the early stages of tumorigenesis.

 The development of cancerous lesions is caused by the acquisition of a series of changes in the DNA that give a cell a selective growth advantage.
15 These changes can be genetic, such as point mutations or deletions, or epigenetic, such as DNA methylation and histone modifications. The main epigenetic modification in mammals is methylation of cytosine residues of CpG dinucleotides, which can alter the expression of genes and is transmitted through cell division. The proximal promoter and first exon
20 regions of 40-50% of human genes contain small (0.5 to several kb) clusters of CpGs, called CpG islands (CGIs) that are protected from methylation [Bird, 1986. Nature 321: 209-13]. This lack of methylation might be a prerequisite for active transcription, as illustrated by two normal exceptions to this situation. Fully methylated CGIs are found only in promoters of
25 silenced alleles for selected imprinted autosomal genes [Li, 1993. Nature 366: 362-5] and multiple silenced genes on the inactivated X-chromosomes of females [Mohandas et al., 1981. Science 211: 393-6; Surani, 1998. Cell 93: 309-12]. Cancer susceptibility may be influenced by differences in stringency of epigenetic control.

30 Tumours are often characterized by an imbalance in cytosine methylation as manifested both by regional hypermethylation of CGIs and

by global hypomethylation [Ehrlich, 2002. *Oncogene* 21: 5400-13; Gama-Sosa et al., 1983. *Nucl. Acids Res* 11: 6883-94; Goelz et al., 1985. *Science* 228: 187-90; Feinberg et al., 1988. *Cancer Res* 48: 1159-61]. The role of aberrant DNA methylation in cancer is well documented. A growing number of cancer genes are being recognized that harbour dense methylation in normally unmethylated promoter CGIs [Jones and Laird, 1999. *Nat Genet* 21: 163-7]. This methylation both marks and plays a key role in an epigenetically mediated loss-of-gene function [Baylin and Herman, 2000. *Trends Genet* 16: 168-74]. Almost half of the tumour-suppressor genes that cause familial cancers through germline mutations can be inactivated in association with promoter hypermethylation in sporadic cancers, including Rb, APC, VHL, p16INK4A, BRCA1, E-cadherin and hMLH1 [Jones and Laird, 1999. *Nat Genet* 21: 163-7; Baylin and Herman, 2000. *Trends Genet* 16: 168-74]. Hypermethylation of hMLH1 causes microsatellite instability (MSI) in sporadic cancers, including colorectal cancer (CRC) [Herman et al., 1998. *Proc Natl Acad Sci U S A* 95: 6870-5; Veigl et al., 1998. *Proc Natl Acad Sci U S A* 95: 8698-702]. In addition to classic tumour-suppressor genes, promoter hypermethylation is being associated with a growing list of other genes that have strongly been implicated in tumorigenesis and for which loss of function appears to be linked primarily with this epigenetic mode of inactivation [Baylin and Herman, 2000. *Trends Genet* 16: 168-74].

Some tumour types have a higher percentage of methylated known CGIs than others: for example the most hypermethylated tumours originate from the gastrointestinal tract (oesophagus, stomach, colon), while significantly less hypermethylation has been reported in e.g. ovarian tumours [Esteller et al., 2001. *Cancer Res* 61: 3225-9]. It has become clear that CGI hypermethylation is not restricted to a few CGIs, but affects multiple loci and exists in a gradual range across cancer cell lines [Paz et al., 2003. *Cancer Res* 63: 1114-21] and primary human tumours [Esteller et al., 2001. *Cancer Res* 61: 3225-9]. Therefore, it is more appropriate to

perform more refined methylation profiling rather than stratification into CGI methylator positive (CIMP+) or negative (CIMP-) phenotypes based on only a few loci [Whitehall et al., 2002. Cancer Res 62: 6011-4 ; Suter et al., 2003. Br J Cancer 88: 413-9 ; Toyota et al., 1999. Proc Natl Acad Sci U S A 96: 8681-6]. Several approaches are now available for profiling CGI methylation in human cancers. Microarray-based approaches have the advantage of being technically simple. They don't require large amounts of DNA, and they can screen thousands of loci in parallel [Ushijima, 2005. Nature Reviews Cancer 5: 223-231]. Huang and co-workers have developed an array-based strategy containing short GC-rich tags for differential methylation hybridization [Huang et al., 1999. Hum Mol Genet 8: 459-70]. Tumour DNA and normal DNA are each digested with methylation-sensitive restriction enzymes, and PCR amplicons derived from each sample are then hybridized to a CGI microarray. Using a panel of around 8000 CGIs, primary tumours could be hierarchically clustered into groups based on their methylation profiles that correlated with histological grade [Yan et al., 2000. Clin Cancer Res 6: 1432-8] and hormone receptor status [Yan et al., 2001. Cancer Res 61: 8375-80] in breast tumours, with progression-free survival in late-stage ovarian cancer [Wei et al., 2002. Clin Cancer Res 8: 2246-52], and with subtypes of cutaneous T cell lymphomas [van Doorn et al., 2005. J Clin Oncol. 23: 3886-96].

Several studies have been performed describing the relationship between CGI methylation and CRC [e.g. Toyota et al., 1999. Proc Natl Acad Sci U S A 96: 8681; Bai et al., 2004. Int J Cancer 112: 846-853; Rashid et al., 2001. Am J Pathol 159: 1129-35]. Using MSP to determine the methylation status of 30 new MINT loci and 3 known tumour-suppressor genes in primary CRCs and adenomas, Toyota and co-workers [Toyota et al., 1999. Proc Natl Acad Sci U S A 96: 8681] found that the majority of CGI methylation events in CRCs are age-related. Virtually all the other methylation events occurred in a distinct subset of CRCs and adenomas that

they termed CGI methylator phenotype positive (CIMP+). Similarly, other studies demonstrated the early and specific involvement of promoter hypermethylation of several tumour-related genes in the colorectal adenoma-carcinoma sequence [Bai et al., 2004. *Int J Cancer* 112: 846-853; Rashid et al., 2001. *Am J Pathol* 159: 1129-35].

The distinction between CIMP+ and CIMP- arises very early in the pathogenesis of CRC [Kondo and Issa, 2004. *Cancer and Metastasis Reviews* 23: 29-39], possibly as early as the aberrant crypt foci. Aberrant methylation also contributes to later stages of colon cancer formation and progression [Kondo and Issa, 2004. *Cancer and Metastasis Reviews* 23: 29-39] and has been related to the serrated pathway [Jass et al., 2000. *Histopathology* 37: 295-301]. The CIMP+ tumours, comprising half of all sporadic CRC, are distinctly characterized by pathology, clinical and molecular genetic features. Genetically, CIMP+ CRCs can be divided into two groups; one including the majority of sporadic MSI-High cancers related to hMLH1 promoter methylation [Toyota et al., 1999. *Proc Natl Acad Sci U S A* 96: 8681], and another with a high incidence of K-ras mutations and a low incidence of p53 mutations [Whitehall et al., 2001. *Cancer Res* 61: 827-30; Toyota et al., 2000. *Proc Natl Acad Sci U S A* 97: 710-5]. Part of the association of CIMP+ with K-ras mutations may be related to silencing of the DNA repair gene MGMT by promoter methylation, which has been reported to increase the incidence of G-A mutations [Esteller et al., 2000. *Cancer Res* 60: 2368-71]. CIMP+ can coexist with APC mutations in sporadic CRC [Hawkins et al., 2002. *Gastroenterology* 122: 1376-87; Gayet et al., 2001. *Oncogene* 20: 5025-32], and is also found in MSI-stable CRC and cell lines [Whitehall et al., 2002. *Cancer Res* 62: 6011-4; Suter et al., 2003. *Br J Cancer* 88: 413-9]. These findings suggest that hypermethylation may be involved in both colorectal tumorigenesis pathways.

Suzuki et al. performed a genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in a human CRC cell line.

Subsequent analysis of paired primary CRC tumours and normal tissues showed an association with hypermethylated 5' CGIs in a tumour-specific manner [Suzuki et al., 2002. Nat Genet 31: 141-9]. These results are promising for future studies into the biological mechanisms and clinical consequences of CGI methylation in CRC. What are still lacking are highly informative methylation markers for cancer detection including its precursor forms and the association of methylation patterns with clinical behaviour of CRC. This requires the methylation profiling of large numbers of well-defined tumours with known clinical outcome.

10

The invention now provides a method for determining whether an individual is suffering from a tumour in a tissue, comprising determining from a sample comprising nucleic acid from said tissue the methylation of a CpG in the genomic region of a receptor protein-tyrosine phosphatase gamma gene and determining from said methylation whether said individual is suffering from a tumour in said tissue. Methylation of a CpG in the genomic region of a receptor protein-tyrosine phosphatase gamma gene (protein tyrosine phosphatase, receptor type, G; *PTPRG*, Ref seq. ID NC_000003) can be used as an early marker for determining whether an individual is suffering from a tumour in a tissue.

20

The term tumour refers to an abnormal increase in the number of cells that has developed in a tissue. A tumour can be benign and not metastasized. Non-limiting examples of a benign tumour are prostatic hyperplasia, cutaneous lymphoid hyperplasia, colorectal polyps or adenomas, and benign ductal or lobular hyperplasia of the breast. While some tumours are not associated with an increased risk of progression into a malignant and/or metastasized tumour, often, however, a tumour is associated with an increased risk of malignant transformation. Non-limiting examples of this type of tumours are C-cell hyperplasia, a premalignant

30

stage in the development of medullary thyroid carcinoma, colorectal polyps or adenomas, and atypical ductal or lobular tumour of the breast, which presence indicate an increased risk of developing cancer.

The term tumour also refers to malignant primary or metastasized tumours typically referred to as neoplasia. Examples thereof include but are not limited to a carcinoma, a sarcoma, a lymphoma, a leukaemia, or a myeloma.

A tumour can be present in any tissue or part of a body, including but not limited to bone, brain, eye, breast, skin, bladder, lung, ureter, thyroid, parathyroid, salivary gland, kidney, prostate, genital system including ovary and testis, endometrium, blood/haematologic system, or in a gastrointestinal tissue. In a preferred embodiment said tumour is an epithelial tumour or at least of epithelial origin.

In a preferred embodiment, a sample comprising nucleic acid from said tissue refers to a sample comprising nucleic acid, preferably desoxyribonucleic acid, from epithelial tumour cells.

In a preferred embodiment, said tumour is present in a gastrointestinal tissue. The gastrointestinal system comprises organs such as mouth, oesophagus, stomach, small and large intestine, anus, liver, bile duct, and pancreas. Tumours of a gastrointestinal tissue comprise a polyp or adenoma, gastrointestinal stromal tumour, gastrointestinal sarcoma, gastrointestinal mesenchymal tumour, leiomyoma, leiomyosarcoma, leiomyoblastoma, and gastrointestinal carcinoma.

A tumour of the gastrointestinal system can be a polyp or adenoma. Polyps are typically of epithelial origin and can be found in tissues comprising a mucous membrane such as colon, small intestine, stomach,

nose, urinary bladder, cervix and uterus. Said polyp can be an inflammatory polyp, a hyperplastic polyp or an adenomatous polyp. Although most polyps themselves are benign, the presence of a polyp is indicative of an increased risk for developing cancer. Adenomatous polyps can be divided into villous,
5 tubular and tubulovillous adenomatous polyps.

A tumour of the gastrointestinal system can also be a carcinoma, a cancer of epithelial tissue that covers or lines surfaces of organs, glands, or body structures. Said carcinoma can be an adenocarcinoma, or an
10 undifferentiated carcinoma.

The transformation of normal gastrointestinal epithelial cells to cancer cells follows a process of molecular and histological changes. The drivers of this process are genetic and epigenetic alterations, leading to
15 growth advantages and expansion of the altered cells. The transformation from a normal epithelial cell to a polyp to a carcinoma occurs over a period of about 10 years, whereby histological changes occur at each step in this process starting with a benign tubular adenoma to an invasive adenocarcinoma

20

In a further preferred embodiment, said tumour is present in colorectal tissue. Colorectal tumour is the fourth most often diagnosed tumour in both male and female, accounting to about 10% of all cancer deaths. Cancer of the colon is highly treatable and often curable by surgery.
25 However, local recurrence or recurrence at a distant site following surgery occurs in about 50% of all cases. Patients of which the tumour had penetrated beyond the bowel wall and/or there was evidence of metastasis to distant organs at the moment of surgery, have a five year survival rate of less than ten percent. In general, early diagnosis and treatment of a

colorectal tumour enhances the survival rate, as this limits the chance of recurrence of the tumour.

A few hereditary diseases are known that increase the risk for developing a tumour of the gastrointestinal system. These diseases include familial adenomatous polyposis, a rare genetic disease in which people develop tumours of the adenomatous type in the colon and often also in the upper intestine; Gardner's syndrome, causing tumours to develop throughout the colon and upper intestine and also in other parts of the body such as skin (sebaceous cysts and lipomas), bone (osteomas) and abdomen (desmoids); *MUTYH*-associated polyposis, a rare autosomal recessive disease caused by mutations in the *MUTYH* gene, the human homologue of the *Escherichia coli* mutY gene; and hereditary nonpolyposis colorectal cancer, causing not only tumours in the colon but also in other organs. Hereditary nonpolyposis colorectal cancer includes Lynch I and Lynch II syndromes. Lynch I syndrome usually leads to the development of a small number of polyps that quickly become malignant. Lynch II syndrome often leads to the development of tumours in the breast, stomach, small intestine, urinary tract and ovaries as well as in the colon. Thus in a preferred embodiment, said individual is suspected or diagnosed as hereditary at risk of developing a tumour of the gastro-intestinal tract.

Methylation of a CpG in the genomic region of a gene often results in transcriptional silencing of the gene through complex effects on transcription factor binding and associated changes in chromatin structure. These effects typically though not necessarily involve methylation of CpGs in promoter/enhancer and/or other transcriptional regulatory sequences. Aberrant methylation may play a role in the transformation process of cells by silencing genes that normally prevent growth of cells. Methylation of CpG can also affect other phenomena in a cell. Sequences that are involved

in these phenomena typically, though not necessarily, reside within a 1 mega base pairs of the genomic sequence of the gene they affect. Following this general theme it is thought that methylation of a CpG in a region more than 1 mega base pairs upstream from a transcription initiation site or more
5 than 1 mega base pairs downstream from a poylyadenylation site are less likely to be genetically linked to allow adequate assessment and/or diagnosis of the risk that said individual has for having and/or developing a tumour of said tissue.

10 In a preferred embodiment, therefore, said genomic region of the receptor protein-tyrosine phosphatase gamma gene is defined herein as a region from 1 mega base pairs upstream from the most upstream transcription initiation site of said gene to 1 mega base pairs downstream from the most distant poylyadenylation site of said gene, more preferred
15 from 100 kilo base pairs upstream from the most upstream transcription initiation site of said gene to 100 kilo base pairs downstream from the most distant poylyadenylation site of said gene, or most preferred from 10 kilo base pairs upstream from the most upstream transcription initiation site of said gene to 10 kilo base pairs downstream from the most distant
20 poylyadenylation site of said gene.

In a preferred embodiment, one end of said region is a region from 1 mega base pairs upstream from the transcription initiation site at chr3:61,522,283 (March 2006 human reference sequence (NCBI Build 36.1)) to 1 mega base
25 pairs downstream from the poylyadenylation site at chr3:62,255,613 of said gene, more preferred from 100 kilo base pairs upstream from the most upstream transcription initiation site of said gene to 100 kilo base pairs downstream from the most distant poylyadenylation site of said gene, or most preferred from 10 kilo base pairs upstream from the most upstream

transcription initiation site of said gene to 10 kilo base pairs downstream from the most distant polyadenylation site of said gene.

In a preferred embodiment, methylation of a CpG that is present in
5 a first intron of said receptor protein-tyrosine phosphatase gamma gene is determined. Methylation of at least one CpG in said first intron was found to be an early marker for the presence of a colorectal tumour in an individual. The genomic region on chromosome 3p14.2 comprises two CpG-rich regions that are present within said intron (see Figure 4). In a
10 preferred method of the invention, methylation of at least one CpG from said genomic region as indicated in Figure 4, preferably at least one CpG selected from CpG1, CpG2, CpG3, CpG4, CpG5, CpG6, CpG7, CpG8, CpG9, CpG 10, CpG11, CpG12, CpG13, CpG14, CpG15, CpG16, CpG17, CpG18, CpG19, CpG 20, CpG21, CpG22, CpG23, CpG24, CpG25, CpG26, CpG27,
15 CpG28, CpG29, CpG 30, CpG31, CpG32, CpG33, CpG34, CpG35, CpG36, and CpG37 as indicated in Figure 4, is determined.

Said genomic region encompasses the region covered by CpG island clone 47B02. Thus in a preferred embodiment, methylation of at least one CpG selected from CpG1, CpG2, CpG3, CpG4, CpG5, CpG6, CpG7, CpG8,
20 CpG9, CpG 10, and CpG11 as indicated in Figure 4 in said CpG island is determined. In a further preferred embodiment, said CpG is selected from CpG7, CpG8, CpG9, and CpG10, as indicated in Figure 4. In a further preferred embodiment, said CpG comprises CpG9 and CpG10, as indicated in Figure 4. Preferably said CpG comprises CpG9.

25

A further preferred method according to the invention comprises determining methylation of at least two of the CpGs indicated in Figure 4, more preferred at least three, more preferred at least four, more preferred at least five, more preferred at least six, more preferred at least seven, more
30 preferred at least eight, more preferred at least nine, more preferred at least

ten, more preferred at least fifteen, more preferred at least twenty, more preferred at least thirty, more preferred at least thirty-seven of the CpGs indicated in Figure 4.

5 A sample according to the invention is preferably isolated from blood, stool, or urine from said individual. For routine testing, it is important that said sample containing nucleic acid can be withdrawn from the individual in a cost-effective and patient-compliant manner. Body secreted, such as blood, stool and urine, can easily be used for these testing.

10 Other preferred samples that can be used include but are not limited to samples comprising skin, hair, saliva, cheek swab, or lung fluid. In a preferred embodiment said sample is a biopsy of an epithelial tissue. In a particularly preferred embodiment said sample is a stool sample. In a preferred embodiment said sample comprises nucleic acid of gastrointestinal

15 cells. It is preferred that said nucleic acid is derived from cells from said tissue. In a further preferred embodiment, the sample comprises nucleic acid of colon cells.

 In a preferred embodiment, the methylation of CpG is determined by

20 comparing to a reference. Said reference can be a sample from an individual of which the presence or absence of a tumour has been previously determined. In a preferred embodiment, said reference is taken from a sample from an individual of which relevant data, comprising position and number of methylated CpG nucleotides have been stored in a database. Said

25 database can be present in an electronic storage device, such as, but not limited to, a computer or a server. It is further preferred that said database comprising said reference can be addressed to compare the position and number of methylated CpG nucleotides with said reference. In a preferred embodiment, said reference comprises at least an unmethylated DNA and/or

30 a fully methylated DNA.

As the markers of the present invention detect early stages of tumorigenesis it is possible to determine whether an individual has such early stages. A method of the invention is therefore very sensitive in that also early stages are detected. When combined with a marker for late stages of tumorigenesis it may be possible to determine whether said individual comprises an early or a late stage tumour. Thus in a preferred embodiment a method of the invention further comprises determining in a sample of said individual the presence or absence of a marker for late stage tumorigenesis. Said tissue can be determined and/or diagnosed to be free of tumour, to comprise early stage tumour or to comprise late stage tumour.

Methylation of CpGs in a sample can be determined using a variety of methods. As also described in the examples herein said methods include but are not limited to differential methylation hybridization, methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA), and methods based on bisulphite modification of DNA including bisulphite sequencing, methylation-specific PCR (MSP) and quantitative variations thereof, methylation-sensitive single nucleotide primer extension (MS-SnuPE), combined bisulphite restriction analysis (COBRA), methylation-sensitive high resolution melting (MS-HRM), array-based methods such as CpG island-specific micro-arrays, and/or mass spectrometry analysis. Genes and/or loci that are affected by aberrant methylation appear to have significant potential to be used as molecular markers for colon tumours as well as for a variety of other tumours.

A preferred method for determining methylation of a CpG according to the invention comprises use of methylation sensitive restriction of the test nucleic acid. Some of the restriction enzymes that are currently available to the artisan are sensitive to methylation and either require methylation for

cleavage of the target nucleic acid or vice versa only cleave the target nucleic acid when it is not methylated. Use of such sensitive restriction enzymes provide test nucleic acid that is cut at the designated target site or not, depending on the methylation state of the target site nucleic acid. The
5 digested nucleic acid can be used directly as a probe or be probed, or it can first be amplified and subsequently used as a probe. In the latter case, when using amplification primers that flank the target site, one will obtain an amplificate if the target site is not cleaved by the enzyme, and vice versa no amplificate when the target site is cleaved. The resulting product, whether
10 or not successfully amplified, can subsequently be used as a probe or be probed. Methylation-dependent restriction of the nucleic acid can be performed by using methylation-sensitive restriction enzymes, including but not limited to BstUI, HpaII and HhaI. A preferred method for detection based on methylation-sensitive restriction enzymes comprises multiplex
15 ligation-dependent probe amplification [Nygren et al., 2005. Nucl. Acids Res 33: e128].

Therefore, in a preferred method according to the invention, methylation of CpG is determined by amplifying the nucleic acid before and
20 after methylation-dependent restriction of said nucleic acid.

Another convenient method is provided by treating the nucleic acid with sodium bisulphite, which converts unmethylated cytosines to uracils, but leaves methylated cytosines unchanged. Methods based on bisulphite-
25 converted DNA include bisulphite sequence analysis [Grunau et al. (2001) Nucl. Acids Res 29: e65], detection of methylation using bead arrays [Bibikova et al., 2006. Genome Res. 16: 383-93], MSP [Herman et al., 1996. Proc Natl Acad Sci USA 93: 9821-6], methylation detection by mass spectrometry [Ehrich et al., 2005. Proc Natl Acad Sci U S A. 102: 15785-90],

Ms-SNuPE [Gonzalzo and Jones, 1997. Nucl Acids Res 25, 2529–2531], and MS-HRM [Wojdacz and Dobrovic, 2007. Nucl Acids Res 35, e41] .

Therefore, in another preferred method according to the invention,
5 methylation of CpG is determined with a method comprising bisulphite modification of said nucleic acid.

In an alternative embodiment, the method comprises labelling of the amplified nucleic acid and hybridization of the labelled nucleic acid to a
10 microarray comprising probes that are able to hybridize to the labelled nucleic acid. The presence and quantity of hybridization signal of the labelled nucleic acid to a probe on the microarray can be determined as is known to a skilled person and is dependent on the label that is used for the nucleic acid. The difference in hybridization signal before and after
15 methylation-dependent restriction of the nucleic acid can be used to determine the methylation of a CpG in said nucleic acid. Alternatively, a difference in hybridization signal is determined between samples from an individual suffering from a tumour in a tissue, or suspected of suffering therefrom, and healthy individuals that are treated with methylation-
20 sensitive restriction enzymes, and/or between samples from an individual suffering from a tumour in a tissue, or suspected of suffering therefrom, and an individual with a tumour.

Especially preferred are methods which allow processing of multiple
25 samples in an economical and time-efficient way, such as MS-MLPA, custom bead arrays, and multiplex PCR methods based on pre-treatment with methylation-sensitive restriction enzymes [Nygren et al., 2005. Nucl Acids Res 33: e128] or based on pre-treatment with bisulphite such as MSP and quantitative derivatives thereof such as quantitative multiplex-MSP [QM-
30 MSP; Fackler et al. 2004. Cancer Research 64, 4442-4452].

In a particularly preferred embodiment, a methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) assay is used to test methylation of specific CpGs in the 3' region of the BSA-validated region in a consecutive CRC validation series (Figure 11). The MS-MLPA
5 assay is robust and can be performed on DNA derived from formalin-fixed paraffin-embedded tissues.

The invention furthermore provides a kit for determining whether a
10 person suffers from a tumour, said kit comprising means for determining the methylation of a CpG present in a genomic region of a receptor protein-tyrosine phosphatase gamma gene in a sample comprising nucleic acid from said person, said genomic region including a region up to 1 mega base pairs upstream from the most upstream transcription initiation site and 1 mega
15 base pairs downstream from the most distal polyadenylation site of said gene, more preferred from 100 kilo base pairs upstream from the most upstream transcription initiation site of said gene to 100 kilo base pairs downstream from the most distant polyadenylation site of said gene, or most preferred from 10 kilo base pairs upstream from the most upstream
20 transcription initiation site of said gene to 10 kilo base pairs downstream from the most distant polyadenylation site of said gene.

Methods to isolate nucleic acid, such as desoxyribonucleic acid, from stool are known in the art and comprise "QIAamp DNA Stool Mini Kit" (Qiagen,
25 the Netherlands) and "PSP® Spin Stool Genomic DNA Purification Kit" (Invitek, Germany).

A kit according to the invention preferably comprises at least two primers that allow amplification of said genomic region comprising said
30 CpG. Amplification can be performed by any method known in the art

including, but not limited to, polymerase chain reaction, strand displacement amplification, nucleic acid sequence-based amplification, rolling circle amplification technology, and transcription-mediated amplification. Each of these amplification methods uses different approaches to achieve the amplification of nucleic acid molecules to amounts that can subsequently be detected. The kit preferably also comprises means for treating the DNA with bisulphite, or means for treating the DNA with a relevant methylation-sensitive restriction enzyme, prior to amplification.

10 In a preferred embodiment, a kit according to the invention provides means for amplifying a CpG that is present in a first intron of said receptor protein-tyrosine phosphatase gamma gene. In this embodiment, a set of primers can be used that allow amplification of said first intron, or at least a part of said intron that comprises a CpG marker. A preferred set of primers is selected from the primers provided in Figure 8. Preferably said set of primers is a set provided in Figure 8C. Particularly preferred is the set indicated by MLPA2 in Figure 8C.

20 The invention also provides the use of a kit according to the invention for determining whether an individual is suffering from a tumour in a tissue. The use of a kit according to the invention provides a cost-effective and patient-compliant way of using an early marker for prognosing or diagnosing an individual for the presence of a tumour.

25 In another embodiment, the invention provides a method for determining whether an individual is suffering from a tumour in a tissue, comprising determining from a sample containing nucleic acid from said tissue an mRNA expression level of said receptor protein-tyrosine phosphatase gamma gene and determining from said expression level whether said individual is suffering from a tumour in said tissue.

30

The mRNA expression level of said receptor protein-tyrosine phosphatase gamma gene can be determined by any method known to a skilled person, including but not limited to Northern blotting and
5 quantitative reverse transcriptase-PCR.

In yet another embodiment, the invention provides a method for determining whether an individual is suffering from a tumour in a tissue, comprising determining from a sample containing protein from said tissue a
10 protein expression level of said receptor protein-tyrosine phosphatase gamma gene and determining from said expression level whether said individual is suffering from a tumour in said tissue.

Said protein expression level of said receptor protein-tyrosine phosphatase gamma gene can be determined by any method known to a skilled person, including but not limited to Western blotting and immunohistochemistry.
15

In a preferred embodiment, said method comprises comparing the determined expression level of said receptor protein-tyrosine phosphatase gamma gene to the expression level of said gene in a reference sample.
20

The invention further provides a method for determining whether an individual is suffering from a tumour in a tissue, comprising determining in
25 a sample comprising nucleic acid from said tissue, the methylation of a binding site for CCCTC-binding factor (zinc finger protein), also called CTCF, in the genomic region of the receptor protein-tyrosine phosphatase gamma gene and determining from said methylation whether said individual is suffering from a tumour in said tissue.

The invention further provides a method for determining whether an individual is suffering from a tumour in a tissue, comprising determining in a sample comprising nucleic acid of the first intron of the receptor protein-tyrosine phosphatase gamma gene from said tissue, whether the CTCF
5 protein, can bind to said nucleic acid of said first intron. Also provided is the use of CTCF protein for determining whether a sample of a tissue of an individual comprises tumor cells. Further provided is a method for determining whether a sample of a tissue of an individual comprises tumor cells comprising determining whether CTCF protein can bind to nucleic acid
10 of the first intron of the receptor protein-tyrosine phosphatase gamma gene from said tissue.

The term "CTCF" refers to a protein involved in insulator activity or to a nucleotide coding for said protein. The gene has the Ref seq. ID NC_000016.
15 The protein CTCF plays among others a role of repressing the insulin-like growth factor 2 gene, by binding to the H-19 imprinting control region (ICR) along with Differentially-methylated Region-1 (DMR1) and MAR3.

Binding of targeting sequence elements by CTCF can block the interaction
20 between enhancers and promoters, therefore limiting the activity of enhancers to certain functional domains. Besides acting as an enhancer blocker, CTCF can also act as a chromatin barrier by preventing the spread of heterochromatin structures.

25 Two independent studies (Xie et al. from MIT and Kim et al. from UCSD) revealed that the human genome contains nearly 15,000 CTCF insulator sites, suggesting a wide-spread role of CTCF in gene regulation (Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES (2007). "Systematic discovery of regulatory motifs in conserved regions of the
30 human genome, including thousands of CTCF insulator sites". Proc. Natl.

Acad. Sci. U.S.A. 104 (17): 7145–50. doi:10.1073/pnas.0701811104. PMID 17442748.

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenko VV, Ren B (2007). "Analysis of the vertebrate
5 insulator protein CTCF-binding sites in the human genome". Cell 128 (6): 1231–45)

It was further revealed that CTCF binding sites act as nucleosome positioning anchors so that, when used to align various genomic signals,
10 multiple flanking nucleosomes can be readily identified (Fu Y, Sinha M, Peterson CL, Weng Z (2008). "The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome". PLoS genetics 4 (7): e1000138. doi:10.1371/journal.pgen.1000138. PMID 18654629)

15

In a preferred embodiment, said CTCF binding site is a CTCF binding site in the region OREG0015647, chr3: 61,525,101-61,525,851 of UCSC March 2006 assembly. In a more preferred embodiment, said CTCF binding site comprises the DNA sequence:

20 ttttctttccctggtgtgtgaggaagcttgagatccaaaatgggactgccaggaaccagccttCG¹tggg
gcttgaccattttCG²tcttctcattttcttttCG³CG⁴cccaCG⁵ctgCG⁶aggtaaatagccctttct
ggtcCG⁷ggagcCG⁸aggggtgtgggaaagggaaaggacagtgggtgggaggCG⁹caggaagaggg
CG¹⁰gtttggttggaagagtcagccCG¹¹agagggagcagcaggctttggagcaaggtaaagt.

In an even more preferred embodiment, said CTCF binding site comprises
25 the DNA sequence: gaaaggacagtgggtgggaggCG⁹caggaagagggCG¹⁰gttt,

Methods of determining whether a protein can bind to DNA are known in the art. A preferred method is by performing chromatin immunoprecipitation with an antibody against said CTCF binding site.

30 Preferably, methylation of said CTCF binding site is determined, preferably

using a restriction enzyme specific for the methylation sensitive HhaI site. In a more preferred embodiment, a CTCF binding site obtained by immunoprecipitation is amplified. In a more preferred embodiment, said CTCF binding site is amplified using a primer according to Figure 8C.

5

The invention further provides a method for determining whether methylation of a binding site for the CTCF protein in the genomic region of the receptor protein-tyrosine phosphatase gamma gene is correlated with the occurrence of a tumor in a tissue sample, said method comprising

10 determining whether methylation of a CpG in said genomic region is correlated with the occurrence of said tumor and determining whether the methylation state of said CpG affects the binding of CTCF to the nucleic acid of the genomic region of the receptor protein-tyrosine phosphatase gamma gene. In a preferred embodiment said CTCF binding is determining

15 in a nucleic of about 50 nucleotides of said genomic region comprising said CpG.

Figure legends

- Figure 1. Methylation profiling by differential methylation hybridization on CpG island clone microarrays detects differentially methylated loci. P-value curve for ANOVA results after testing for differences in methylation ratios between three histology groups: carcinoma, adenoma, normal. The red dotted line indicates the multiple testing corrected p-value cutoff of 0.0001. At this cutoff, 20 loci were selected, the most significant one clone 47B02.
- Figure 2. The 47B02 locus is hypermethylated in tumours. Variance plot of the ANOVA in Figure 1, showing the log10 ratios of the grouped samples for the *PTPRG* intron 1 locus. The log ratios of tumour and adenoma samples compared to the tumour cell line reference panel are close to 0, compared to a negative log ratio for the normals.
- Figure 3. Methylation profiles of 20 selected loci. Trend plot of the 'top-20' differentially methylated loci (see Figure 1), showing the log10 ratios in the normal, adenoma and carcinoma hybridizations. As all of the loci have a lower log10 ratio in the normal group (n=5) in comparison to the carcinoma group (n=17) one may conclude that these loci are hypermethylated in carcinomas. The log10 ratios of the adenoma group (n=2) are all in the same order as the normal group, except for the *PTPRG* intron 1 clone (dark blue line).
- Figure 4. Genomic sequence encompassing CpG island clone 47B02 located at chr3:61,524,993-61,525,363 (UCSC assembly of March 2006) (shaded), and downstream CpG-rich region (chr3:61,525,384-61,525,620). All CpGs are indicated in red. The locations of one pair of primers for bisulfite sequence analysis are underlined. Note that the primer sequences are different, since they are based on DNA sequences after bisulfite

- modification, see Figure 8A (BSA). The location of one MS-MLPA assay (MLPA2, see Figure 8C) is double underlined. The 11 CpGs in the 47B02 sequence are in capitals and numbered 1-11. Restriction sites used for differential methylation hybridization amplicon generation are boxed: ccgg, HpaII; cgcg BstUI. The last 264 bp of the 47B02 sequence overlap with the first part of OREG0015647 (chr3:61,525,101-61,525,851). The start of the overlap is indicated by the * located four basepairs downstream of cg^{min1} and continues 231 bp beyond the end of the sequence given here.
- (B) UCSC Genome Browser view of part of chromosome 3p (chr3: 61,524,969-61,525,930) showing part of intron 1 of *PTPRG*. Indicated are the relative size and location of CpG island clone 47B02 (blue), a small CpG island (light green), and regulatory element OREG0015647 from the OregAnno database (dark green; Kim et al. 2007 Cell 128:1231-45).
- Figure 5. Colon tumour specific methylation of *PTPRG* intron 1 locus. The dot diagram indicates direct bisulfite sequencing results of 10 CpG dinucleotides in 18 colon tumours (red bar) and 19 paired normal colon samples (green bar), using the BSA primers indicated in Figure 8A. Black dot: methylated CpG; White dot: unmethylated CpG. Grey dot: sequence not readable. The black arrow heads indicate adenoma samples. On top are indicated the locations of the two methylation-sensitive restriction enzymes used in the amplicon generation for differential methylation hybridization to CpG island microarrays.
- Figure 6. Clonal bisulfite sequence analysis of *PTPRG* intron 1 locus confirmed direct bisulfite sequence results. The dot diagram indicates clonal bisulfite sequencing results of 10 CpG dinucleotides in (A) adenoma tID180 and (B) carcinoma tID127. N indicates a cloned allele from the normal tissue DNA, T for the tumour DNA. Black dot: methylated CpG; White dot: unmethylated CpG. Grey dot: sequence not readable. On top are indicated

the locations of the two methylation-sensitive restriction enzymes used in the amplicon generation for differential methylation hybridization to CpG island microarrays.

- 5 Figure 7. Bisulfite mass spectrometry analysis of *PTPRG* intron 1 locus. Representative example shown for carcinoma tID184 (80% tumour cells) and paired normal tissue. For the MS primers, the specific primer sequences were identical to the BSA primers and a tail was added to allow mass spectrometry application (see Figure 8A). RNase cleaved fragments are
10 scored based on the shift in mass between a methylated fragment and an unmethylated fragment. The y-axis shows the percentage of methylated fragments in comparison with unmethylated ones. The values are the means and standard errors of three independent measurements. Fragments 1-3 contain CpGs 1-3 respectively, fragment 4 contains CpGs 4 and 5, fragment
15 5 contains CpGs 6 and 7, and fragments 6-9 contain one CpG each, i.e. CpGs 8-11.

- Figure 8. Primers for methylation detection of the *PTPRG* intron 1 locus. A. Bisulfite-sequence analysis (BSA) and methylation-specific mass
20 spectrometry (MS). B. Methylation-specific PCR (MSP) on bisulfite modified DNA. C. Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA).

- Figure 9. Specificity and sensitivity of 9 individual CpGs in the 47B02
25 region measured by direct bisulfite sequence analysis (see Figure 5). Sensitivity indicates the percentage of tumours with methylation of the 47B02 genomic region. Specificity indicates the percentage of unmethylated normals.

Figure 10. Distribution 0% and 100% methylated DNA controls. Cutoffs were based on controls included in ten individual experiments, including the fully methylated DNA (Chemicon-Millipore) and the unmethylated human semen DNA. Tested samples falling within three standard deviations of the mean of unmethylated and methylated control ratios were typed accordingly. Samples with ratios in between these boundaries were typed partially methylated.

Figure 11. *PTPRG* intron 1 CpG9 methylation detected by MS-MLPA.

10 Methylation frequency of *PTPRG* intron 1 CpG9 in carcinomas (Ca(T)), advanced adenomas (AA(T)) and corresponding normal epithelial tissue (CA(N) and AA(N)) as well as in precursor lesions hyperplastic polyps (HP), serrated adenoma (SA), early adenoma (EA). The number of tumors typed as methylated (dark), partially methylated (striped) and unmethylated

15 (white) in the MS-MLPA assay is indicated.

Figure 12. Sensitivity and specificity for *PTPRG* intron 1 CpG9 methylation in carcinomas and advanced adenomas.

Examples

Example 1

DNA samples

Anonymised tumour and normal colon mucosa biopsies were obtained and
5 fresh-frozen. For isolation of DNA, pathologist-checked macrodissected
(trimmed) frozen sections were used to minimize the percentage of normal
epithelium and stromal cells. To control for patient-dependent methylation,
we used normal epithelium from the same individuals as controls, where
available. About 20 sections of 30 µm yielded at least 30-50 µg of DNA,
10 which was sufficient for microarray hybridization and confirmation with
alternative methods. Control DNA samples representing fully methylated
(CpGenome universal methylated) and unmethylated (CpGenome universal
unmethylated) DNA were obtained from Chemicon/Millipore. Colorectal
carcinoma cell lines SW48, RKO, SW480, Caco2, SW837, and LS411 were
15 obtained from the American Type Culture Collection (Manassas, US) and
cultured according to the manufacturer's instructions. DNA was isolated
using standard protocols [Isola et al., 1994. Am J Pathol 145: 1301-1308].

CpG island microarrays

20 We obtained a copy of the 8600 CGI clone library from Dr. T.H. Huang
(Center for Integrative Cancer Biology, The Ohio State University,
Columbus, Ohio), based on a library originally generated by the Sanger
Centre from affinity-purified in vitro methylated MseI-digested DNA
fragments [Cross et al., 1994. Nat Genet 6: 236-44]. The library was
25 sequenced at the Toronto Microarray Facility. CGI clone inserts were
amplified using vector-based primers essentially as described [Knijnenburg
et al., 2005. Am J Med Gen 132A: 36-40; Yan et al., 2002. Methods 27: 162-
9]. Three promoter regions of tumour-related genes were added, *RASSF1*
(forward TAATTGCCAATGAGGAAAGGGGAAGT, reverse
30 CCGCAACCGTTAAGACTGAAACGT), *MLH1* (forward

- CCATGCACTGGTATACAAAGTCCC, reverse
GATGCGCTGTACATGCCTCT), and *MSH2* (forward
GCCTTGCAGCTGAGTAAACACAGAAAG, reverse
GTGCCTCCGCACTGGAGAGGCTGCTCA). The PCR products were filtered
5 and spotted in 150 mM sodium phosphate buffer pH 8.5, 0.0002% sarkosyl
onto CodeLink (GE Healthcare) slides using an OmniGrid arrayer (Genomic
Solutions) as described [Knijnenburg et al., 2005. Am J Med Gen 132A: 36-
40].
- 10 Differential methylation hybridization
- The detection of differential methylation was based on methylation-
sensitive restriction of MseI-digested, linker-ligated genomic DNA, followed
by linker-mediated PCR amplification [Huang et al., 1999. Hum Mol Genet
8: 459-70]. Genomic fragments containing aberrantly methylated sites are
15 protected from the digestion and can be amplified by linker-PCR, whereas
the same fragments containing the unmethylated sites are cut and cannot
be amplified. Sequential digestion with two methylation-sensitive
endonucleases, HpaII and BstUI, enhances detection of CpG loci with
extensive methylation in tumours and reduces the risk of incomplete
20 digestion [Yan et al., 2001. Cancer Res 61: 8375-80]. Low amplification (20
cycles) was used to prevent overamplification of unrestricted repetitive
sequences in the ligated DNA and yet yield sufficient PCR products for
single or low-copy number CGI loci. Cy5-labeled amplicons, representing
pools of methylated DNA fragments derived from tumours and normal
25 controls, were co-hybridized with a Cy3-labeled common reference amplicon
to the CGI microarrays. In analogy to gene expression profiling, we chose a
common reference design to allow comparison of methylation across all
tumour and normal samples. The common reference consisted of a pool of
six CRC cell lines (SW48, RKO, SW480, Caco2, SW837, LS411). Detection

was done on a G2565BA scanner (Agilent Technologies), and image analysis using GenePix6.0 (Molecular Devices).

Analysis of CGI microarrays

Preprocessing and statistical analysis was performed in Rosetta Resolver (Rosetta Biosoftware) using custom R plug-ins for error-weighted ANOVA with multiple testing correction [Hochberg and Benjamini, 1990. Stat Med. 9: 811-8].

Bisulphite validation methods

Patient DNA samples (500 ng) were converted using the EZ DNA methylation Gold bisulfite kit (Zymo Research). Primers for bisulfite sequence analysis (BSA) of the *PTPRG* intron 1 locus were designed using MethPrimer [Li and Dahiya, 2002. Bioinformatics 18: 1427-31]. Left primer 5'-GATTTAAAATGGGATTGTTAGGGAAT-3' and right primer 5'-CTTACTCCAAAACCTACTACTCCCTCT-3'. For clonal BSA, up to ten colonies of cloned bisulfite PCR product were sequenced for both normal and tumour. The resulting PCR product (224 bp) was sequenced using the right primer. The same specific primer sequences were extended with T7 RNA polymerase promoter sequences for base-specific cleavage and mass spectrometry analysis [Ehrich et al., 2005. Proc Natl Acad Sci U S A 102: 15785-15790].

RESULTS

CpG island methylation profiling of colorectal tumours

CpG island microarray profiling was used for the high-throughput analysis of methylation status of 8.6K CpG islands in 17 right-sided carcinomatous, 2 adenomatous and 5 corresponding normal colonic epithelium samples. The microarray data were tested for differential methylation between tumours (including both carcinomas and adenomas) and normal samples using error-weighted analysis of variance. We identified 22 loci with a false discovery rate (FDR) < 1%. In addition, an ANOVA for the three histology groups was performed, although the adenoma group contained only two samples. In this analysis, we identified 20 loci with a very stringent FDR < 0.01% (Figure 1). The most significant CpG island clone in both analyses was 47B02, which mapped to the first intron of the *PTPRG* gene. The log ratios of tumour and adenoma samples compared to the CRC cell line reference panel are close to 0, indicating comparable methylation levels in the primary tumours and the CRC cell lines. The normal samples showed negative log ratios, indicating that the intron 1 *PTPRG* CGI was hypermethylated in both carcinomas and the adenomas compared to normal colon (Figure 2). This finding indicates that the methylation of the *PTPRG* locus could be an early event in tumourigenesis. The remaining selected loci were hypermethylated in a high percentage of carcinomas, but not in the tested adenomas and normal colon samples (Figure 3).

Example 2

Validation and extension of array data using bisulphite sequence analysis

The *PTPRG* intron 1 methylation status was validated using direct bisulphite sequence analysis (BSA). Part of the 47B02 clone sequence was used to design BSA primers (Figure 4). Additional right-sided adenoma (2) and carcinoma (7) samples were included as well as two left-sided tumours

(one adenoma, one carcinoma). For all but one tumour, the methylation status of paired normal samples was determined (Figure 5). The selected amplified region contains two methylation-sensitive restriction sites that were used in the amplicon generation for the differential methylation hybridization. BSA gives a resolution of single CpGs; CpGs 2 through 10 could be evaluated using this method. All of the 18 tumour samples showed methylation of the region, while one carcinoma showed partial methylation. In contrast, the normal samples were mostly unmethylated, with six samples showing partial methylation of one to five CpGs. CpGs 7-10 showed the best distinction between tumour and normal in this set of samples. The tumour-specific methylation frequency of this locus is very high; based on the current data set we could estimate a sensitivity of 94-100% to detect methylation in adenoma/carcinoma tissue and a specificity of 94-100% for CpGs 7-10 (see Figure 9). Therefore, the methylation microarray results were confirmed and extended to additional proximal and distal adenomas and carcinomas.

To study the methylation of the *PTPRG* intron 1 locus at the single chromosome level, clonal bisulphite analysis was performed on four pairs of tumour-normal samples. Clonal BSA confirmed the direct BSA results, and showed that in the partially methylated normal samples at most three out of ten alleles were methylated for several CpGs (Figure 6). Since the percentage of tumour cells varied between 30% (T74) and 80% in the tumour biopsies, as expected we find some alleles unmethylated in the sequenced tumour clones.

Example 3

Quantitative assessment of *PTPRG* intron 1 methylation

To further confirm our findings, we used a quantitative high-throughput analysis of *PTPRG* intron 1 methylation patterns by base-specific cleavage

and mass spectrometry [Ehrich et al., 2005. Proc Natl Acad Sci U S A 102: 15785-15790] on nine previously analyzed tumour-normal pairs. A representative example result for tID184 illustrates tumour-specific methylation of all fragments, covering 11 CpG dinucleotides (Figure 7).

5 Fragment 6 gave overall low quality results because of its small size. The other eight tumour-normal pairs also confirmed tumour specific hypermethylation of all the fragments present in the amplified *PTPRG* locus. A percentage between 35-80% methylation of tumour fragments was found, which correlated well with the estimated percentage of tumour cells
10 in the analyzed tissue.

Therefore, we identified and confirmed tumour-specific methylation, most specifically for four CpGs in the 47B02 clone locus in the first intron of *PTPRG*, in 94% of tested carcinomas and adenomas from different locations in the colon and rectum. Initial analysis of expression of the main isoform of
15 *PTPRG* mRNA by quantitative RT-PCR did not show a consistent effect, however a small reduction in expression associated with methylation of the intron 1 locus as well as decreased expression of alternative transcripts cannot be ruled out.

20 Example 4

METHODS

MS-MLPA probe design and assay

Custom MS-MLPA probes for the *PTPRG* locus were designed in primer3 [Rozen and Skaletsky 2000. Methods Mol Biol 132:365-386] and included
25 CpG 9 and 10 (as numbered in Figure 4A) in the sequence investigated by BSA . Probes used are: PTPRG_L: 5'- GAAAGGACAGTGGTGGGAGGC -3' (Tm 63.9°C) and PTPRG_R: 5'-GCAGGGAAGAGGGCGGTT -3' (Tm 63.36°C), genomic region Chr3: 61525269-61525308 (UCSC assembly: March 2006).

As a control we have used a *BRCA2* probeset from the MRC-Holland SALSA MS-MLPA KIT ME001B Tumor suppressor-1 kit: *BRCA2_L*: 5'-GGCCATGGAATCTGCTGAACAAAA – 3' and *BRCA2_R*: 5'-GGAACAAGGTTTATCAAGGGATGTCACAACCGTGTGGAAGTTGCG – 3' ;
 5 genomic region Chr13: 31851549 – 31851617 (UCSC assembly: March 2006). Fragment analysis was performed on an ABI 3130 (Applied Biosystems, Foster City, US). MLPA reagents were obtained from MRC-Holland, Amsterdam, The Netherlands (EK1 kit; www.mlpa.com). Approximately 50 ng of genomic DNA in 5 µl of water was used as input.
 10 The MS-MPLA was performed as described by Nygren et al [Nygren et al. 2005. Nucleic Acids Res 33:e128]. Negative (human semen DNA) and 100% methylated DNA controls (CpGenome Universal methylated DNA, Chemicon(Millipore), Billerica, MA, USA), were included every run to assess HhaI cleavage.

15

MS-MLPA analysis

Fragment run analysis was performed in Genemapper (Applied Biosystems, Foster City, US). Peak heights were exported and processed in Excel.
 20 *PTPRG* peak heights were normalized by division with the *BRCA2* peak heights of the same run. Subsequently, the ratio of the HhaI-digested reaction was divided by the ratio of the undigested reaction providing one ratio per sample. Ten individual measurements of the unmethylated DNA and fully methylated control DNAs provided a ratio distribution for each.
 25 Tested samples falling within three standard deviations of the mean of the unmethylated and methylated reference samples were typed accordingly. Samples containing ratios between these standard deviation boundaries were typed partially methylated (Figure 10).

30

RESULTS

MS-MLPA validation of *PTPRG* intron 1 methylation

We developed a methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) assay to test methylation of specific CpGs in the 3' region of the BSA-validated region in a consecutive CRC validation series (Figure 11). The MS-MLPA assay is robust and can be performed on DNA derived from formalin-fixed paraffin-embedded tissues. Of the 67 tested carcinoma samples 91% showed (partial) methylation of the targeted CpG dinucleotides, while 95.8% of the corresponding normal mucosa samples was unmethylated (n=48, Figure 12). Even higher levels of sensitivity and specificity were obtained for advanced adenomas (Figure 12). A few lesions preceding the adenoma/carcinoma stages were tested for which no corresponding normal tissue was available. Sensitivities of 83.3% for hyperplastic polyps (n=6) and 66.7% for serrated adenomas (n=12) were obtained (Figure 11). *PTPRG* intron 1 CpG9 was also methylated in sporadic mismatch repair deficient colon tumors due to *MLH1* promoter methylation and surprisingly in all 11 colon cancers tested from patients suffering from the Lynch syndrome with germline mutations in one of the MMR genes.

20

Genomic locus overlaps with regulatory feature

Overlapping with the 47B02 sequence, a regulatory feature has been annotated in the human genome (OREG0015647, chr3: 61,525,101-61,525,851 of UCSC March 2006 assembly) based on chromatin immunoprecipitation-on-chip studies for binding sites of the vertebrate insulator protein CTCF [Kim et al. 2007. Cell 128: 1231-45]. Most of the experimentally discovered CTCF binding sites in this study are located far from the transcriptional start sites, with their distribution strongly correlated with genes. Also, CTCF binding sites are largely invariant across different cell types [Kim et al. 2007. Cell 128: 1231-45].

30

Recent studies have identified CTCF to be the vertebrate insulator protein [Bell et al. 1999. *Cell* 98: 387–396]. Insulator elements affect gene expression by preventing the spread of heterochromatin and restricting transcriptional enhancers from activation of unrelated promoters. So far, CTCF remains as
5 the only major protein implicated in establishment of insulators in vertebrates [Felsenfeld et al. 2004. *Cold Spring Harb. Symp. Quant. Biol.* 69: 245–250], including those involved in regulation of gene imprinting and monoallelic gene expression [Fedoriw et al. 2004. *Science* 303: 238–240.; Ling et al. 2006. *Science* 312: 269–272], as well as in X chromosome
10 inactivation and in the escape from X-linked inactivation [Filippova et al. 2005. *Dev. Cell* 8: 31–42; Lee, 2003. *Curr. Biol.* 13: R242–R254]. The mechanism of insulator function remains unclear. One model proposes that insulators, by formation of special chromatin structures, compete for enhancer-bound
15 activators, preventing the activation of downstream promoters [Bulger and Groudine, 1999. *Genes Dev.* 13: 2465–2477]. Alternatively, insulators may facilitate the formation of loops, for example, via attachment of chromosomal regions to the nuclear membrane [Yusufzai et al. 2004. *Mol. Cell* 13: 291–298], keeping the intermediate regions exposed for only local
20 interactions between enhancers and promoters. Consistent with this model, it was recently shown that CTCF could mediate long-range chromosomal interactions in mammalian cells, providing a possible mechanism by which insulators establish regulatory domains [Kurukuti et al. 2006. *Proc. Natl. Acad. Sci. USA* 103, 10684–10689; Ling et al. 2006. *Science* 312: 269–272; Yusufzai et al. 2004. *Mol. Cell* 13: 291–298]. Methylation of several CTCF
25 binding sites was shown to abolish CTCF binding [Bell et al. 2000. *Nature* 405: 482–485; Filippova et al. 2005. *Dev. Cell* 8: 31–42; Hark et al. 2000. *Nature* 405, 486–489; Kanduri et al. 2000. *Curr. Biol.* 10: 853–856; Mukhopadhyay et al. 2004. *Genome Res.* 14: 1594–1602].

We performed chromatin immunoprecipitation with an antibody against human CTCF on normal human fibroblast chromatin. In the bound fraction, we amplified the MS-MLPA product MLPA2 (double underline in Figure 4A). The fragment was not amplified after HhaI digestion, indicating that

5 CpG9 is unmethylated in normal fibroblasts. Therefore, in the overlapping region between clone 47B02 and OREG0015647, we experimentally confirmed binding of CTCF protein to a 39 bp region covered by the MLPA2 primers in fibroblasts. These 39 bp lie within the most tumor-specific methylated region.

Claims

1. A method for determining whether an individual is suffering from a tumour in a tissue, comprising determining from a sample comprising nucleic acid from said tissue the methylation of a CpG in the genomic region of the receptor protein-tyrosine phosphatase gamma gene and determining from said methylation whether said individual is suffering from a tumour in said tissue.
5
2. A method according to claim 1, wherein said tissue is a gastrointestinal tissue.
10
3. A method according to claim 2, wherein said tissue is colorectal tissue.
4. A method according to any one of claims 1-3, wherein one end of said region is at 1 mega base pairs upstream from the transcription initiation site at chr3:61,522,283 (March 2006 human reference sequence (NCBI Build 36.1)) and a second end is at 1 mega base pairs downstream from the poylyadenylation site at chr3:62,255,613 of said gene.
15
20
5. Method according to any of claims 1-4, wherein said CpG is present in the first intron of said receptor protein-tyrosine phosphatase gamma gene.
- 25 6. Method according to any of the previous claims, wherein said CpG comprises at least one of the CpGs indicated in Figure 4.

7. Method according to claim 6, wherein said CpG is selected from CpG1, CpG2, CpG3, CpG4, CpG5, CpG6, CpG7, CpG8, CpG9, CpG10, and CpG11 indicated in Figure 4.
- 5 8. Method according to any of the previous claims, whereby said CpG comprises a combination of at least two of the CpGs indicated in Figure 4.
9. Method according to any of the previous claims, wherein the
10 sample containing nucleic acid is isolated from blood, stool, or urine from said individual.
10. Method according to any of the previous claims, wherein said sample comprises nucleic acid from gastrointestinal cells.
- 15 11. Method according to any one of claims 1-10, wherein the methylation of CpG is determined by comparing to a reference.
12. Method according to any of the previous claims, wherein the
20 methylation of CpG is determined by amplifying the nucleic acid before and after methylation-dependent restriction of said nucleic acid.
13. Method according to any of claims 1-11, wherein methylation of
25 CpG is determined with a method comprising bisulphite modification of said nucleic acid.
14. Kit for determining whether a person suffers from a tumour, said kit comprising means for determining the methylation of a CpG
30 present in a genomic region of the receptor protein-tyrosine

phosphatase gamma gene, and including a region up to 1 mega base pairs upstream from the transcription initiation site and 1 mega base pairs downstream from the polyadenylation site of said gene, in a sample comprising nucleic acid from said person.

5

15. Kit according to claim 14, further comprising means for amplifying a qCpG that is present in the first intron of the receptor protein-tyrosine phosphatase gamma gene.

10

16. Kit according to claim 14 or claim 15, comprising a set of primers as indicated in Figure 8.

17. Use of a kit according to any one of claims 14-16, for determining whether an individual is suffering from a tumour in a tissue.

15

18. A method for determining whether an individual is suffering from a tumour in a tissue, comprising determining from a sample containing nucleic acid from said tissue a mRNA expression level of said receptor protein-tyrosine phosphatase gamma gene and determining from said expression level whether said individual is suffering from a tumour in said tissue.

20

19. A method for determining whether an individual is suffering from a tumour in a tissue, comprising determining from a sample containing protein from said tissue a protein expression level of said receptor protein-tyrosine phosphatase gamma gene product and determining from said expression level whether said individual is suffering from a tumour in said tissue.

25

20. Method according to claim 18 or 19, further comprising comparing the determined expression level of said receptor protein-tyrosine phosphatase gamma gene to the expression level of said gene in a reference sample.
- 5
21. A method for determining whether an individual is suffering from a tumour in a tissue, comprising determining in a sample comprising nucleic acid of the genomic region of the receptor protein-tyrosine phosphatase gamma gene from said tissue, whether the CCCTC-binding factor (zinc finger protein), also called CTCF protein, can bind to said nucleic acid of said genomic region of the receptor protein-tyrosine phosphatase gamma gene.
- 10
22. A method for determining whether an individual is suffering from a tumour in a tissue, comprising determining in a sample comprising nucleic acid of the first intron of the receptor protein-tyrosine phosphatase gamma gene from said tissue, whether the CCCTC-binding factor (zinc finger protein), also called CTCF protein, can bind to said nucleic acid of said first intron.
- 15
23. Use of the CTCF protein for determining whether a sample of a tissue of an individual comprises tumor cells.
- 20
24. A method for determining whether a sample of a tissue of an individual comprises tumor cells comprising determining whether CTCF protein can bind to nucleic acid of the first intron of the receptor protein-tyrosine phosphatase gamma gene from said tissue.
- 25

Figure 1

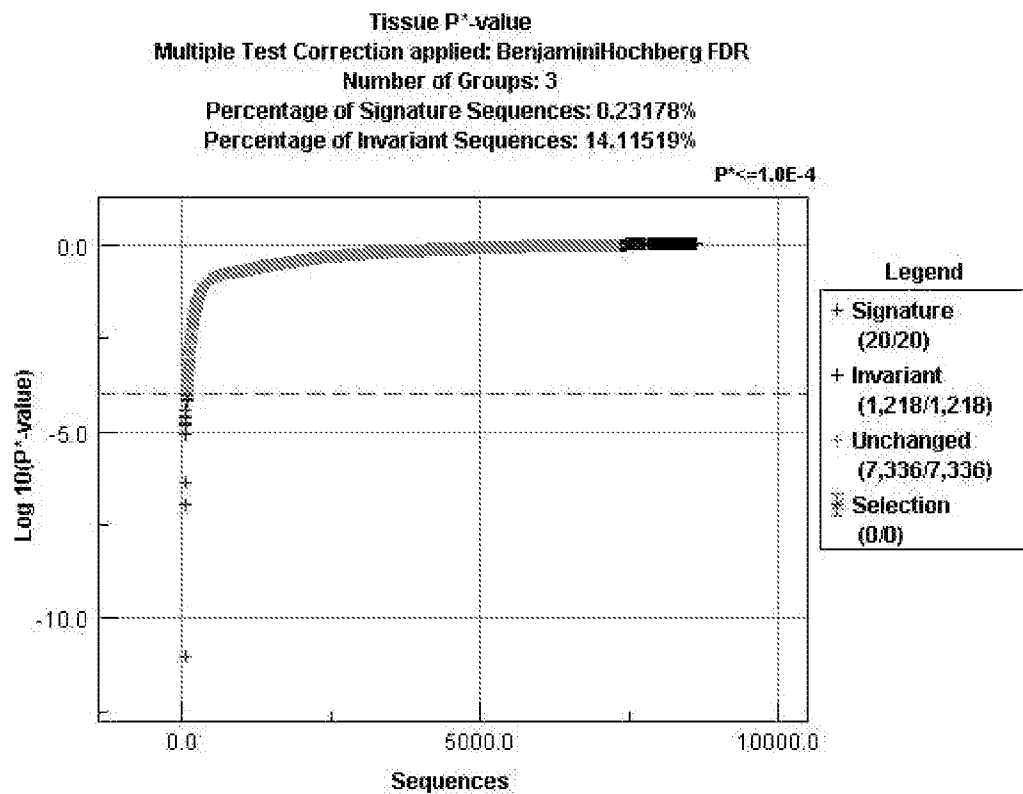


Figure 2

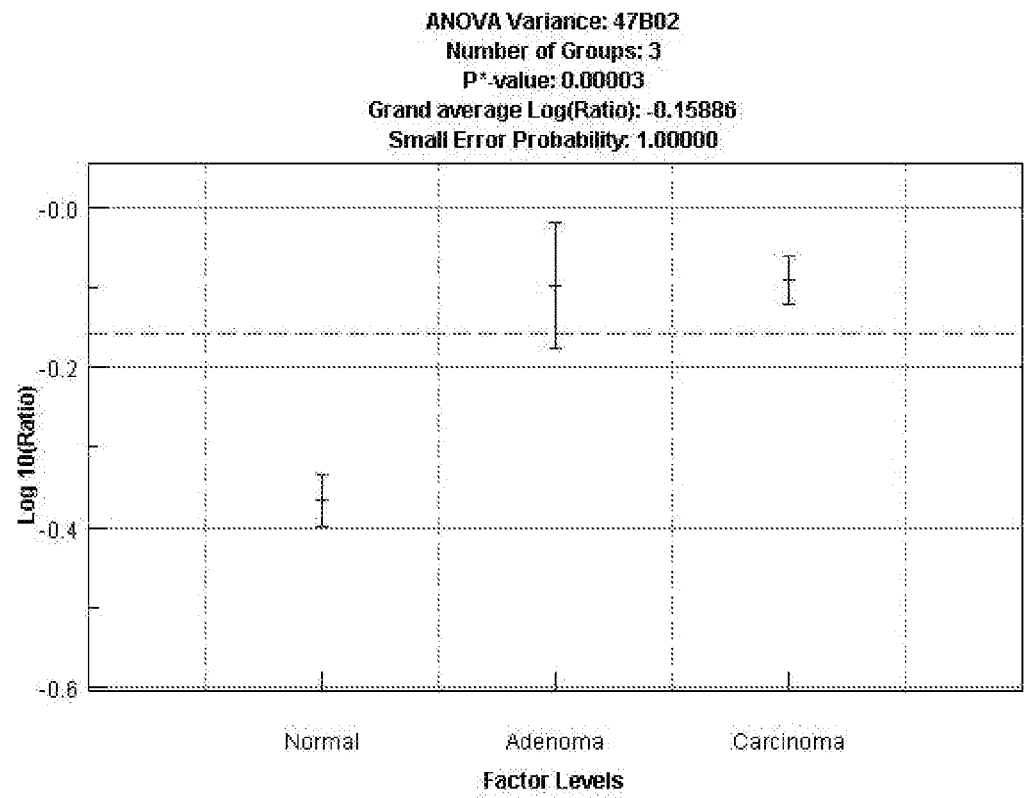


Figure 3

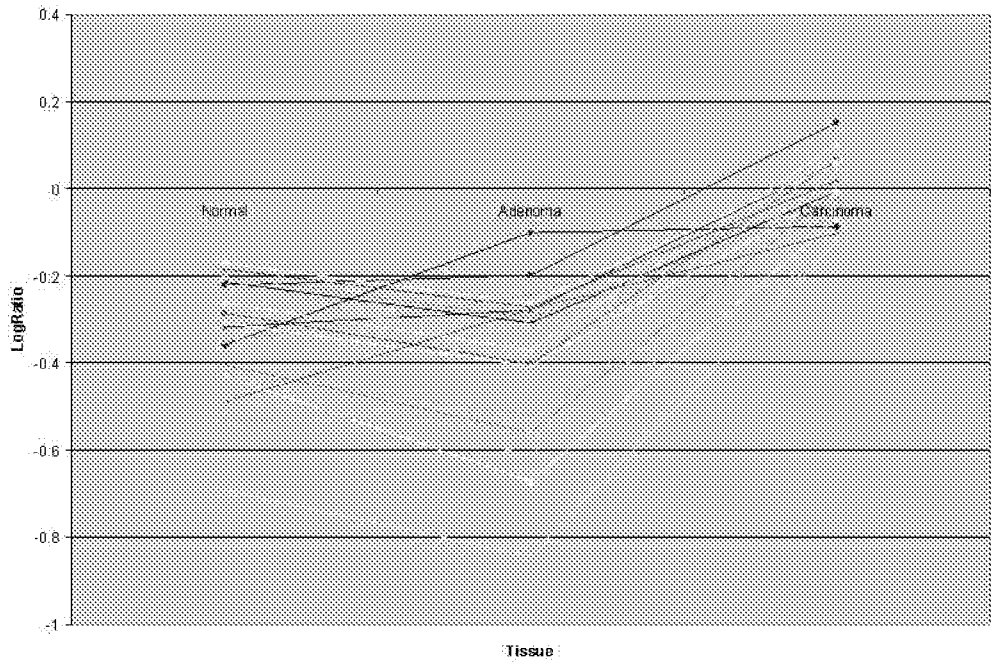


Figure 4A

atgggtgcagtgagggtg¹ tggagggctgacctgcaaatagaggaatttccaattccacttctgtctcttctg²
 g³ctg⁴gggatgcagtcagtgaggctgac⁵ tgg⁶ttttcttttccctgggtgtgtgaggaagcttgagatc⁷
 caaaatgggactgccagggaaccagccttCG tggggcttgaccattttCG tcttctcattttcttttCG CG c⁸
 ccaCG ctgCG aggtaaatagccccctttctgtctCG gtagcCG aggggtgtgggaanagggaaggaca⁹
 gtgggtgggaggCG cagggaagaggCG gttgggtttggaaaagtgcagcctCG agagggagcagcagg¹⁰
 ctltggagcaaggtaaagttaaataatcagagctctgggagacg¹¹ctggcttttctg¹²ttttcog¹³aggttgcg¹⁴
 g¹⁵acg¹⁶ccg¹⁷ctggacctcagggggcg¹⁸cccccg¹⁹agccactggtggggctctgccacctccacactggtcg²⁰
 cctcg²¹gccacctccacg²²cctcagggatggggcg²³cg²⁴cg²⁵tgcccg²⁶ggttcccaggtctctggggtgcagac²⁷
 g²⁸cg²⁹ccttggcg³⁰cg³¹aggcg³²gcg³³gcctcg³⁴gcctggcg³⁵atgctgctctggttttctcg³⁶aaaccg³⁷c³⁸
 tg

Figure 4B

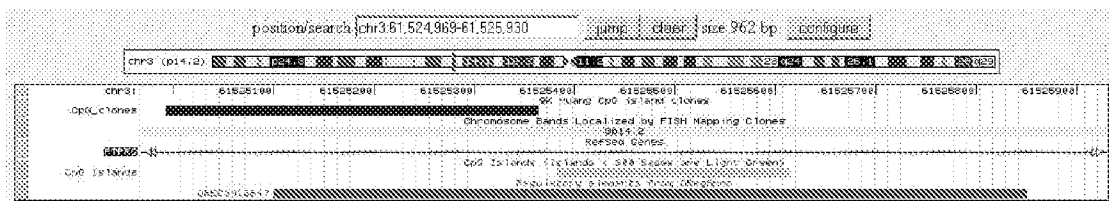


Figure 5

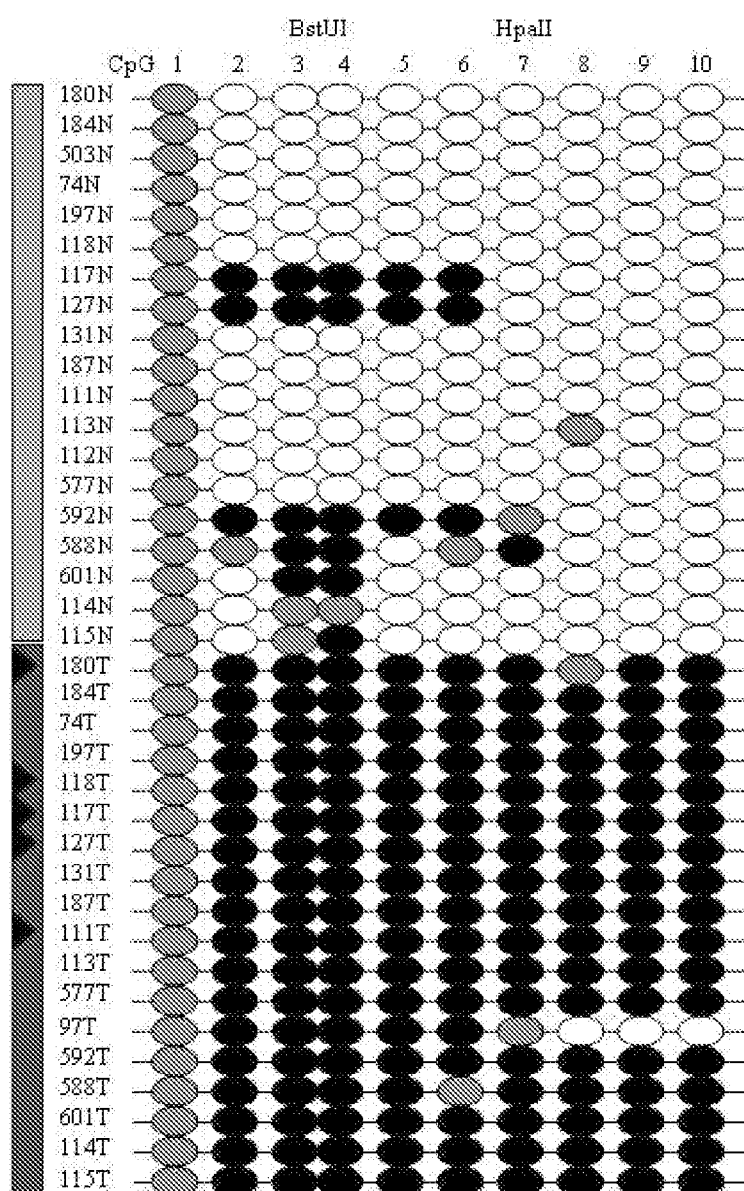


Figure 6

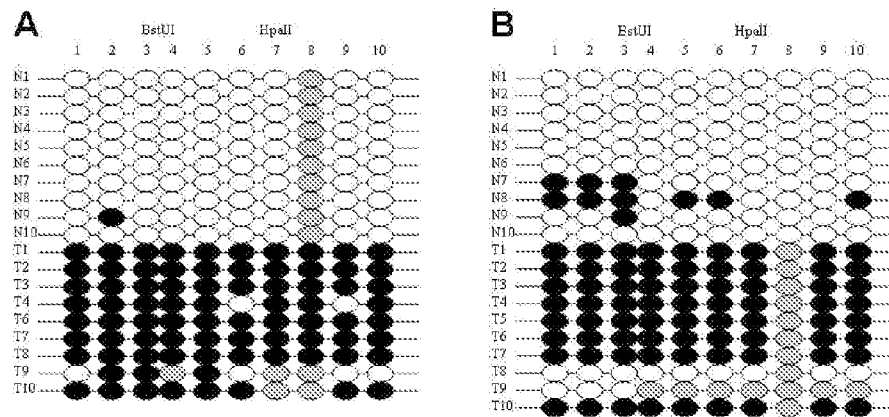


Figure 7

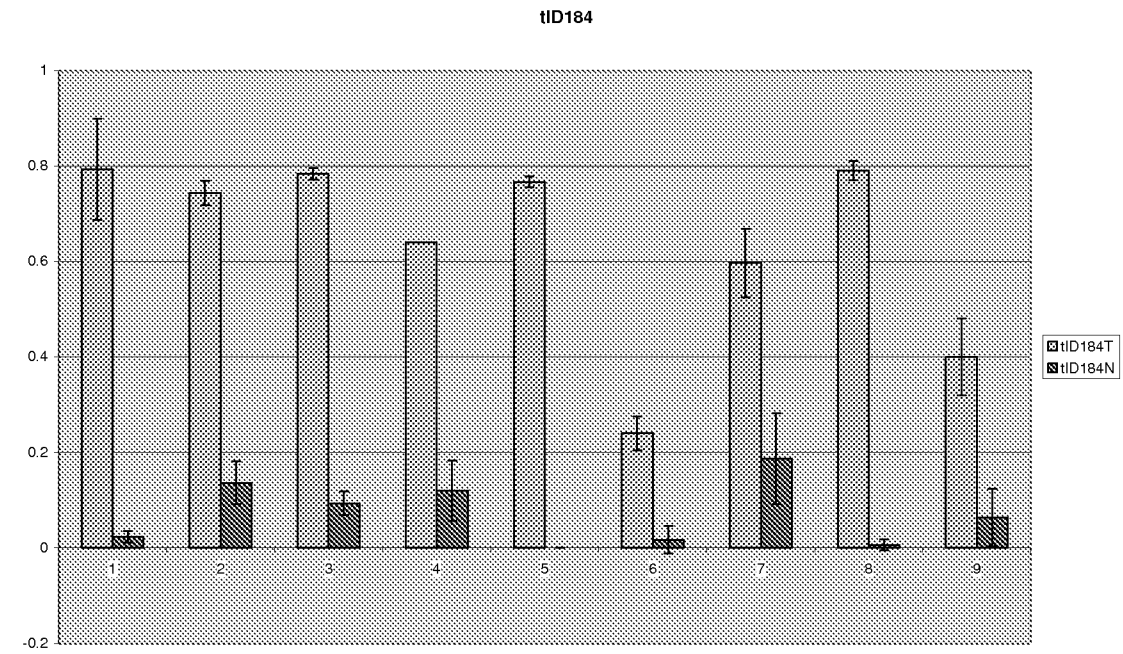


Figure 8A

Primer name	Primer length	Primer sequence (5'-3')	Product size	Product location ¹
BSA				
jb_47B02bsaL	26	GATTTAAAATGGGATTGTTAGGGAAT	224	61,525,133 – 61,525,357
jb_47B02bsaR	27	CTTACTCCAAAACCTACTACTCCCTCT		
BSA-SHORT1				
PTPRG_BSAS1_FW ²	24	TCTCAAACCTTCCTCACACACCAAA	129	61,525,012 – 61,525,134
PTPRG_BSAS1_RV ²	24	GTGGAGGGTTGTTTTGTAAATT		
BSA-SHORT2				
PTPRG2_BSAS2_FW	22	AGAGGGAGTAGTAGGTTTTGGA	136	61,525,330 – 61,525,465
PTPRG2_BSAS2_RV	20	AAAAACCCACCAATAACTC		
MS				
PTPRG_FW_MS ³	36	AGGAAGAGAG GATTTAAAATGGGATTGTTAGGGAAT	265	61,525,133 – 61,525,357
PTPRG_RV_MS ³	58	CAGTAATACGACTCACTATAGGGAGAAGGCT CTTACTCCAAAACCTACTACTCCCTCT		

¹ Base locations refer to chromosome 3, UCSC assembly March 2006.

² These primers are designed to the minus strand.

³ Tail sequences are in bold.

Figure 8B

Primer name	Primer length	Primer sequence (5'-3') ¹	Product size	Product location ²
MSP1 methylated allele				
PTPRG_MSP1_M_Left	23	GGGATGTAGTTAGTGGGTTGATC	155	61,525,073 – 61,525,228
PTPRG_MSP1_M_Right	22	TATTTACCTCGCAAACGTAAACG		
MSP1 unmethylated allele				
PTPRG_MSP1_U_Left	23	GGGATGTAGTTAGTGGGTTGATT	161	61,525,073 – 61,525,234
PTPRG_MSP1_U_Right	29	AAAAACTATTTACCTCACAAACATAAACAC		
MSP2 methylated allele				
PTPRG_MSP2_M_Left	19	TTTTTCGGGTTTACGTTGC	136	61,525,200-61,525,335
PTPRG_MSP2_M_Right	22	CCCTCTCGAACTACACTTTTCC		
MSP2 unmethylated allele				
PTPRG_MSP2_U_Left	22	TTTTTTTTTGTGTTTATGTTGT	140	61,525,197-61,525,338
PTPRG_MSP2_U_Right	23	TCCCTCTCAAACCTACACTTTTCC		
MSP3 methylated allele				
PTPRG_MSP3_M_Left	25	AAAATATTAGAGTTTTGGGAGACGT	139	61,526,365 – 61,525,504
PTPRG_MSP3_M_Right	22	ACGTAAAAATAACCGAAACCGA		
MSP3 unmethylated allele				
PTPRG_MSP3_U_Left	25	AAAATATTAGAGTTTTGGGAGATGT	141	61,526,365 – 61,525,506
PTPRG_MSP3_U_Right	25	AAACATAAAAAATAACCAAAACCAAC		

¹ CpG dinucleotides are indicated in red.² Base locations refer to chromosome 3, UCSC assembly March 2006.

Figure 8C

Primer name	Primer length	Primer sequence (5'-3') ¹	Product size	Ligation location ²
MLPA1				
PTPRG_MLPA1_Left	45	GGGTTCCCTAAGGGTTGGA GAGGTAAATAGCCCCTTCCTGGTCC	86	61,525,244 HpaI site CpG7
PTPRG_MLPA1_Right	41	P-GGGAGCCGAGGGGTGTGT TCTAGATTGGATCTTGCTGGCGC		
MLPA2				
PTPRG_MLPA2_Left	40	GGGTTCCCTAAGGGTTGGA GAAAGGACAGTGGTGGGAGGC	82	61,525,289 HhaI site CpG9
PTPRG_MLPA2_Right	42	P-GCAGGGAAGAGGGCGGTTT TCTAGATTGGATCTTGCTGGCGC		
MLPA3				
PTPRG_MLPA3_Left	48	GGGTTCCCTAAGGGTTGGA CCATTTTCGTCTTCTCATTTTCTTTTCGC	90	61,525,207 HhaI CpG4
PTPRG_MLPA3_Right	42	P-GCCACGCTGCGAGGTAAA TCTAGATTGGATCTTGCTGGCGC		
MLPA4				
PTPRG_MLPA4_Left	50	GGGTTCCCTAAGGGTTGGA TAATTTCCAATTCCACTTCTGTCCTTCTTGC	94	61,525,068 HhaI CpG -3
PTPRG_MLPA4_Right	44	P-GCTCGGGGATGCAGTCAGTGT TCTAGATTGGATCTTGCTGGCGC		

¹ The tail sequences (in bold) are used for amplification using universal primers:

MLPA_F (FAM-5'-GGGTTCCCTAAGGGTTGGA-3') and

MLPA_R (5'-GCGCCAGCAAGATCCAATCTAGA-3').

² Base locations refer to chromosome 3, UCSC assembly March 2006.

Figure 9

CpG	Sensitivity % methylated tumors/all tumors		Specificity % unmethylated normals/all normals	
2	100%	(18/18)	83%	(15/18)
3	100%	(18/18)	71%	(12/17)
4	100%	(18/18)	67%	(12/18)
5	100%	(18/18)	84%	(16/19)
6	100%	(17/17)	83%	(15/18)
7	100%	(17/17)	94%	(17/18)
8	94%	(16/17)	100%	(18/18)
9	94%	(17/18)	100%	(19/19)
10	94%	(17/18)	100%	(19/19)

Figure 10

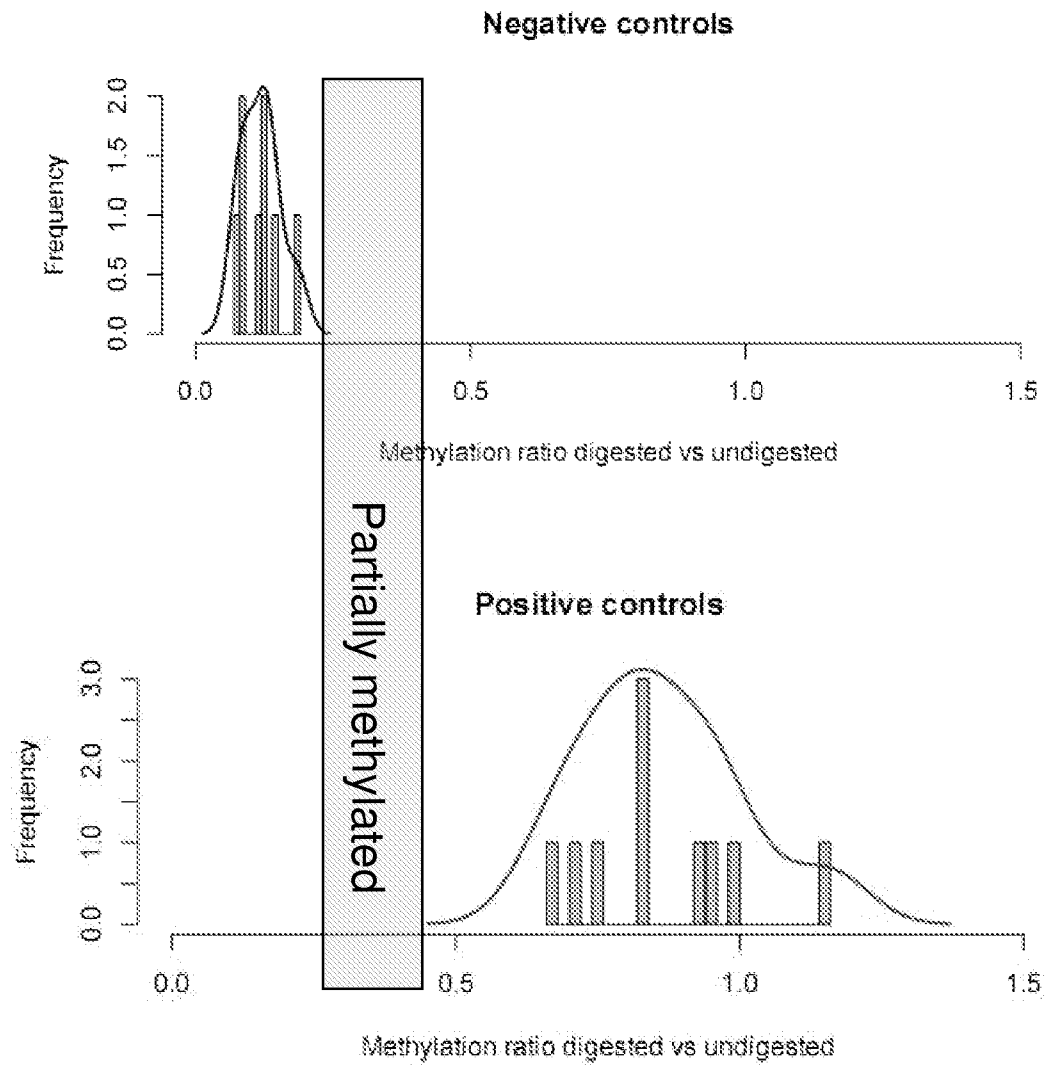


Figure 11

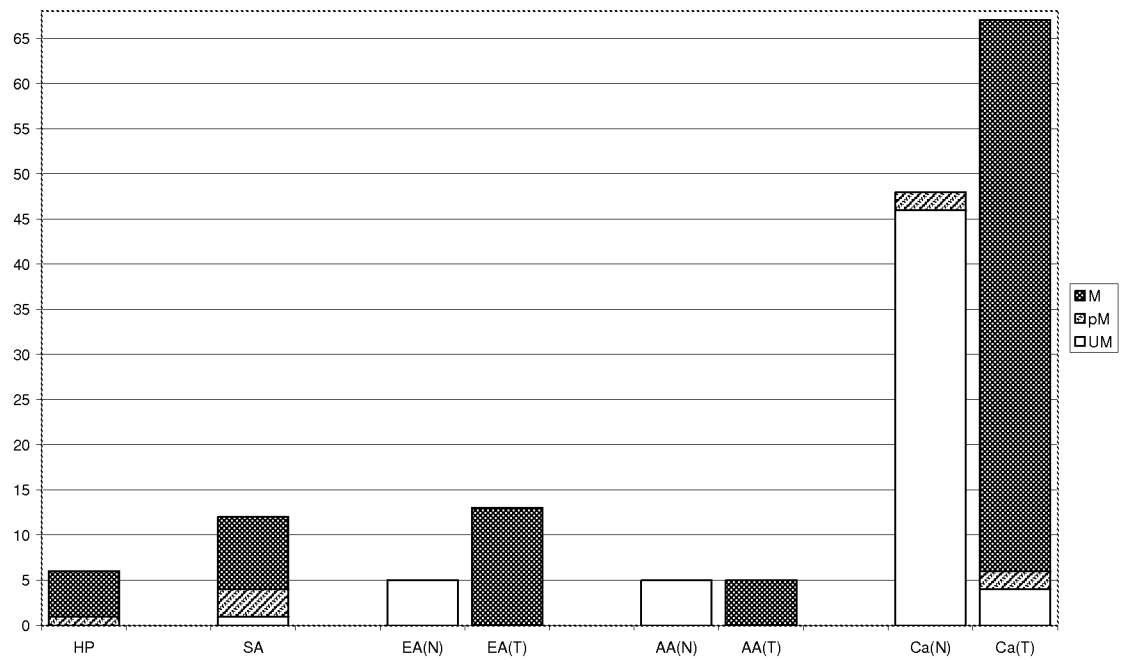


Figure 12

CpG 9	Sensitivity % methylated tumors /total tumors	Specificity % unmethylated normals /total normals
Advanced adenomas	100% (18/18)	100% (10/10)
Carcinomas	91% (61/67)	95.8% (46/48)